

LLECTUAL PROPERTY ORGANIZATION International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5:

C12N 1/24, 15/00, C07H 21/00

(11) International Publication Number:

WO 92/06176

A1

(43) International Publication Date:

16 April 1992 (16.04.92)

(21) International Application Number:

PCT/US91/07141

(22) International Filing Date: 27 September 1991 (27.09.91)

(30) Priority data: 590,664

28 September 1990 (28.09.90) US

(71) Applicant: IXSYS, INC. [US/US]; 3550 General Atomics Court, Suite L103, San Diego, CA 92121 (US).

(72) Inventor: HUSE, William, D.; 471 Avenida Primavera, Del Mar, CA 92014 (US).

(74) Agents: CAMPBELL, Cathryn et al.; Pretty, Schroeder, Brueggemann & Clark, 444 South Flower Street, Suite 2000, Los Angeles, CA 90071 (US).

(81) Designated States: AT (European patent), AU, BB, BE (European patent), BF (OAPI patent), BG, BJ (OAPI pa'ent), BR, CA, CF (OAPI patent), CG (OAPI patent), CH (European patent), CI (OAPI patent), CM (OAPI patent), CS, DE (European patent), DK (European patent), ES (European patent), FI, FR (European patent), GA (OAPI patent), GB (European patent), GN (OAPI patent), GR (European patent), HU, IT (European patent), JP, KP, KR, LK, LU (European patent), MC, MG, ML (OAPI patent), MN, MR (OAPI patent), MW, NL (European patent), NO, PL, RO, SD, SE (European patent), SN (OAPI patent), SU+,TD (OAPI patent), TG (OAPI patent).

Published

With international search report.

(54) Title: SURFACE EXPRESSION LIBRARIES OF RANDOMIZED PEPTIDES

(57) Abstract

A composition of matter comprising a plurality of procaryotic cells containing a diverse population of expressible oligonucleotides operationally linked to expression elements, said expressible oligonucleotides having a desirable bias of random codon sequences.

+ DESIGNATIONS OF "SU"

Any designation of "SU" has effect in the Russian Federation. It is not yet known whether any such designation has effect in other States of the former Soviet Union.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT Austria ES Spain AU Australia FI Finland	MG ML MN MR	Madagascar Mali Mongolia
BB Barbados GA Gabon BB Belgium GB United Kingdom BB Burkina Faso GB United Kingdom BB Burkina Faso GR Guinea BJ Benin HU Hungary BR Brazil HU Hungary TT Italy CA Canada JP Japan CF Central African Republic Of Korea CH Switzerland KR Republic of Korea CI Côte d'Ivoire LI Liechtenstein CM Cameroon LK Sri Lanka DE Germany MC Mouseo BR France GR Greace Hu Hungary TT taly Japan KP Democratic People's Republic of Korea LI Liechtenstein LK Sri Lanka LU Lissembourg MC Mouseo	MW NL NO FL SD SE SU TO TC US	Mauritania Malawi Netherlands Norway Poland Romania Sudan Sweden Senegal Soviet Union Chad Togo United States of America

30

SURFACE EXPRESSION LIBRARIES OF RANDOMIZED PEPTIDES

BACKGROUND OF THE INVENTION

This invention relates generally to methods for synthesizing and expressing oligonucleotides and, more particularly, to methods for expressing oligonucleotides having random codon sequences.

Oligonucleotide synthesis proceeds via linear coupling 10 of individual monomers in a stepwise reaction. reactions are generally performed on a solid phase support by first coupling the 3' end of the first monomer to the The second monomer is added to the 5' end of the 15 first monomer in a condensation reaction to yield a dinucleotide coupled to the solid support. At the end of each coupling reaction, the by-products and unreacted, free monomers are washed away so that the starting material for the next round of synthesis is the pure oligonucleotide In this reaction scheme, the 20 attached to the support. stepwise addition of individual monomers to a single, growing end of a oligonucleotide ensures accurate synthesis of the desired sequence. Moreover, unwanted side reactions eliminated, such as the condensation of 25 oligonucleotides, resulting in high product yields.

In some instances, it is desired that synthetic oligonucleotides have random nucleotide sequences. This result can be accomplished by adding equal proportions of all four nucleotides in the monomer coupling reactions, leading to the random incorporation of all nucleotides and yielding a population of oligonucleotides with random sequences. Since all possible combinations of nucleotide sequences are represented within the population, all possible codon triplets will also be represented. If the

objective is ultimately to generate random peptide products, this approach has a severe limitation because the random codons synthesized will bias the amino acids incorporated during translation of the DNA by the cell into polypeptides.

The bias is due to the redundancy of the genetic code.

There are four nucleotide monomers which leads to sixtyfour possible triplet codons. With only twenty amino acids
to specify, many of the amino acids are encoded by multiple
codons. Therefore, a population of oligonucleotides
synthesized by sequential addition of monomers from a
random population will not encode peptides whose amino acid
sequence represents all possible combinations of the twenty
different amino acids in equal proportions. That is, the
frequency of amino acids incorporated into polypeptides
will be biased toward those amino acids which are specified
by multiple codons.

To alleviate amino acid bias due to the redundancy of the genetic code, the oligonucleotides can be synthesized from nucleotide triplets. Here, a triplet coding for each of the twenty amino acids is synthesized from individual monomers. Once synthesized, the triplets are used in the coupling reactions instead of individual monomers. By mixing equal proportions of the triplets, synthesis of oligonucleotides with random codons can be accomplished. However, the cost of synthesis from such triplets far exceeds that of synthesis from individual monomers because triplets are not commercially available.

Amino acid bias can be reduced, however, by synthesizing the degenerate codon sequence NNK where N is a mixture of all four nucleotides and K is a mixture guanine and thymine nucleotides. Each position within an oligonucleotide having this codon sequence will contain a total of 32 codons (12 encoding amino acids being

represented once, 5 represented twice, 3 represented three times and one codon being a stop codon). Oligonucleotides expressed with such degenerate codon sequences will produce peptide products whose sequences are biased toward those amino acids being represented more than once. Thus, populations of peptides whose sequences are completely random cannot be obtained from oligonucleotides synthesized from degenerate sequences.

There thus exists a need for a method to express oligonucleotides having a fully random or desirably biased sequence which alleviates genetic redundancy. The present invention satisfies these needs and provides additional advantages as well.

SUMMARY OF THE INVENTION

The invention provides a plurality of procaryotic cells containing a diverse population of expressible oligonucleotides operationally linked to expression elements, the expressible oligonucleotides having a desirable bias of random codon sequences.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic drawing for synthesizing oligonucleotides from nucleotide monomers with random tuplets at each position using twenty reaction vessels.

Figure 2 is a schematic drawing for synthesizing oligonucleotides from nucleotide monomers with random tuplets at each position using ten reaction vessels.

Figure 3 is a schematic diagram of the two vectors used for sublibrary and library production from precursor oligonucleotide portions. M13IX22 (Figure 3A) is the v ctor used to clone the anti-sense precursor portions

.4

(hatched box). The single-headed arrow represents the Lac p/o expression sequences and the double-headed arrow represents the portion of M13IX22 which is to be combined The amber stop codon for biological selection and relevant restriction sites are also shown. M13IX42 (Figure 3B) is the vector used to clone the sense precursor portions (open box). Thick lines represent the pseudo-wild type (Ψ gVIII) and wild type (gVIII) gene VIII sequences. The double-headed arrow represents the portion of M13IX42 which is to be combined with M13IX22. The two amber stop codons and relevant restriction sites are also shown. Figure 3C shows the joining of vector population from sublibraries to form the functional surface expression vector M13IX. Figure 3D shows the generation of a surface 15 expression library in a non-suppressor strain and the The phage are used to infect a production of phage. suppressor strain (Figure 3E) for surface expression and screening of the library.

Figure 4 is a schematic diagram of the vector used for generation of surface expression libraries from random oligonucleotide populations (M13IX30). The symbols are as described for Figure 3.

Figure 5 is the nucleotide sequence of M13IX42 (SEQ ID NO: 1).

25 Figure 6 is the nucleotide sequence of M13IX22 (SEQ ID No: 2).

Figure 7 is the nucleotide sequence of M13IX30 (SEQ ID NO: 3).

Figure 8 is the nucleotide sequence of M13ED03 (SEQ ID 30 NO: 4).

Figure 9 is the nucleotide sequence of M13IX421 (SEQ

ID NO: 5).

Figure 10 is the nucleotide sequence of M13ED04 (SEQ ID NO: 6).

DETAILED DESCRIPTION OF THE INVENTION

This invention is directed to a simple and inexpensive method for synthesizing and expressing oligonucleotides having a desirable bias of random codons using individual The method is advantageous in that individual monomers are used instead of triplets and by synthesizing 10 only a non-degenerate subset of all triplets, codon Thus, the oligonucleotides redundancy is alleviated. synthesized represent a large proportion of possible random obtained. can be sequences which oligonucleotides can be expressed, for example, on the 15 surface of filamentous bacteriophage in a form which does not alter phage viability or impose biological selections against certain peptide sequences. The oligonucleotides produced are therefore useful for generating an unlimited number of pharmacological and research products.

invention entails embodiment, the In one 20 sequential coupling of monomers to produce oligonucleotides The coupling with a desirable bias of random codons. reactions for the randomization of twenty codons which specify the amino acids of the genetic code are performed in ten different reaction vessels. Each reaction vessel contains a support on which the monomers for two different codons are coupled in three sequential reactions. One of the reactions couples an equal mixture of two monomers such that the final product has two different codon sequences. The codons are randomized by removing the supports from the reaction vessels and mixing them to produc a single batch of supports containing all twenty codons at a particular position. Synthesis at the next codon position proceeds by

equally dividing the mixed batch of supports into ten reaction vessels as before and sequentially coupling the monomers for each pair of codons. The supports are again mixed to randomize the codons at the position just synthesized. The cycle of coupling, mixing and dividing continues until the desired number of codon positions have been randomized. After the last position has been randomized, the oligonucleotides with random codons are cleaved from the support. The random oligonucleotides can then be expressed, for example, on the surface of filamentous bacteriophage as gene VIII-peptide fusion proteins. Alternative genes can be used as well.

In its broadest form, the invention provides a diverse population of synthetic oligonucleotides contained in vectors so as to be expressible in cells. Such populations of diverse oligonucleotides can be fully random at one or more codon sites or can be fully defined at one or more site, so long as at least one site the codons are randomly variable. The populations of oligonucleotides can be expressed as fusion products in combination with surface proteins of filamentous bacteriophage, such as M13, as with gene VIII. The vectors can be transfected into a plurality of cells, such as the procaryote E. coli.

The diverse population of oligonucleotides can be formed by randomly combining first and second precursor populations, each precursor population having a desirable bias of random codon sequences. Methods of synthesizing and expressing the diverse population of expressible oligonucleotides are also provided.

In a preferred embodiment, two populations of random oligonucleotides are synthesized. The oligonucleotides within each population encode a portion of the final oligonucleotide which is to be expressed. Oligonucleotides within one population encode the carboxy terminal portion

of the expressed oligonucleotides. These oligonucleotides are cloned in frame with a gene VIII (gVIII) sequence so . that translation of the sequence produces peptide fusion The second population of oligonucleotides are 5 cloned into a separate vector. Fach oligonucleotide within this population encodes the anti-sense of the amino terminal portion of the expressed oligonucleotides. This vector also contains the elements necessary for expression. The two vectors containing the random oligonucleotides are 10 combined such that the two precursor oligonucleotide portions are joined together at random to form a population larger oligonucleotides derived from two smaller portions. The vectors contain selectable markers to ensure the two together joining efficiency in maximum 15 oligonucleotide populations. A mechanism also exists to control the expression of gVIII-peptide fusion proteins during library construction and screening.

As used herein, the term "monomer" or "nucleotide monomer" refers to individual nucleotides used in the 20 chemical synthesis of oligonucleotides. Monomers that can be used include both the ribo- and deoxyribo- forms of each of the five standard nucleotides (derived from the bases adenine (A or dA, respectively), guanine (G or dG), cytosine (C or dC), thymine (T) and uracil (U)). Derivatives and precursors of bases such as inosine which are capable of supporting polypeptide biosynthesis are also Also included are chemically included as monomers. modified nucleotides, for example, one having a reversible blocking agent attached to any of the positions on the 30 purine or pyrimidine bases, the ribose or deoxyribose sugar or the phosphate or hydroxyl moieties of the monomer. Such blocking groups include, for example, dimethoxytrityl, benzoyl, isobutyryl, beta-cyano thyl and diisopropylamine groups, and are used to protect hydroxyls, exocyclic amines and ph sphate moieties. Other blocking agents can also be 35 used and are known to one skilled in the art.

As used herein, the term "tuplet" refers to a group of elements of a definable size. The elements of a tuplet as used herein are nucleotide monomers. For example, a tuplet can be a dinucleotide, a trinucleotide or can also be four or more nucleotides.

As used herein, the term "codon" or "triplet" refers
to a tuplet consisting of three adjacent nucleotide
monomers which specify one of the twenty naturally
occurring amino acids found in polypeptide biosynthesis.

The term also includes nonsense, or stop, codons which do
not specify any amino acid.

"Random codons" or "randomized codons," as used herein, refers to more than one codon at a position within a collection of oligonucleotides. The number of different codons can be from two to twenty at any particular position. "Randomized oligonucleotides," as used herein, refers to a collection of oligonucleotides with random codons at one or more positions. "Random codon sequences" as used herein means that more than one codon position within a randomized oligonucleotide contains random codons. For example, if randomized oligonucleotides are six nucleotides in length (i.e., two codons) and both the first and second codon positions are randomized to encode all twenty amino acids, then a population of oligonucleotides having random codon sequences with every possible combination of the twenty triplets in the first and second position makes up the above population of randomized The number of possible codon oligonucleotides. Likewise, if randomized is 20². combinations 30 oligonucleotides of fifteen nucleotides in length are synthesized which have random codon sequences at all positions encoding all twenty amino acids, then all triplets coding for each of the twenty amino acids will be found in equal proportions at very position. 35 population constituting the randomized oligonucleotides will contain 20¹⁵ different possible species of oligonucleotides. "Random tuplets," or "randomized tuplets" are defined analogously.

As used herein, the term "bias" refers to a preference. It is understood that there can be degrees of preference or bias toward codon sequences which encode particular amino acids. For example, an oligonucleotide whose codon sequences do not preferably encode particular amino acids is unbiased and therefore completely random.

The oligonucleotide codon sequences can also be biased toward predetermined codon sequences or codon frequencies and while still diverse and random, will exhibit codon sequences biased toward a defined, or preferred, sequence.

"A desirable bias of random codon sequences" as used herein, refers to the predetermined degree of bias which can be selected from totally random to essentially, but not totally, defined (or preferred). There must be at least one codon position which is variable, however.

As used herein, the term "support" refers to a solid
phase material for attaching monomers for chemical
synthesis. Such support is usually composed of materials
such as beads of control pore glass but can be other
materials known to one skilled in the art. The term is
also meant to include one or more monomers coupled to the
support for additional oligonucleotide synthesis reactions.

As used herein, the terms "coupling" or "condensing" refers to the chemical reactions for attaching one monomer to a second monomer or to a solid support. Such reactions are known to one skilled in the art and are typically performed on an automated DNA synthesizer such as a MilliGen/Biosearch Cyclone Plus Synthesizer using procedures recommended by the manufacturer. "Sequentially coupling" as used her in, refers to the stepwise addition of monomers.

A method of synthesizing oligonucleotides having random tuplets using individual monomers is described. The method consists of several steps, the first being synthesis of a nucleotide tuplet for each tuplet to be randomized. As described here and below, a nucleotide triplet (i.e., a codon) will be used as a specific example of a tuplet. Any size tuplet will work using the methods disclosed herein, and one skilled in the art would know how to use the methods to randomize tuplets of any size.

amino acids is desired at a position, then twenty different codons are synthesized. Likewise, if randomization of only ten codons at a particular position is desired then those ten codons are synthesized. Randomization of codons from two to sixty-four can be accomplished by synthesizing each desired triplet. Preferably, randomization of from two to twenty codons is used for any one position because of the redundancy of the genetic code. The codons selected at one position do not have to be the same codons selected at the next position. Additionally, the sense or anti-sense sequence oligonucleotide can be synthesized. The process therefore provides for randomization of any desired codon position with any number of codons.

by coupling the first monomer of each codon to separate supports. The supports for the synthesis of each codon can, for example, be contained in different reaction vessels such that one reaction vessel corresponds to the monomer coupling reactions for one codon. As will be used here and below, if twenty codons are to be randomized, then twenty reaction vessels can be used in independent coupling reactions for the first twenty monomers of each codon. Synthesis proceeds by sequentially coupling the second monomer of each codon to the first monomer to produce a dimer, followed by coupling the third monomer for each

codon to each of the above-synthesized dimers to produce a trimer (Figure 1, step 1, where M_1 , M_2 and M_3 represent the first, second and third monomer, respectively, for each codon to be randomized).

first codons from synthesis of the Following individual monomers, the randomization is achieved by mixing the supports from all twenty reaction vessels which contain the individual codons to be randomized. The solid phase support can be removed from its vessel and mixed to achieve a random distribution of all codon species within the population (Figure 1, step 2). The mixed population of supports, constituting all codon species, are then redistributed into twenty independent reaction vessels The resultant vessels are all (Figure 1, step 3). identical and contain equal portions of all twenty codons 15 coupled to a solid phase support.

For randomization of the second position codon, synthesis of twenty additional codons is performed in each of the twenty reaction vessels produced in step 3 as the condensing substrates of step 1 (Figure 1, step 4). Steps 1 and 4 are therefore equivalent except that step 4 uses the supports produced by the previous synthesis cycle (steps 1 through 3) for codon synthesis whereas step 1 is the initial synthesis of the first codon in the oligonucleotide. The supports resulting from step 4 will each have two codons attached to them (i.e., a hexanucleotide) with the codon at the first position being any one of twenty possible codons (i.e., random) and the codon at the second position being one of the twenty possible codons.

For randomization of the codon at the second position and synthesis of the third position codon, st ps 2 thr ugh 4 are again repeated. This process yields in each vessel a three codon oligonucleotide (i.e., 9 nucleotides) with

codon positions 1 and 2 randomized and position three Steps 2 containing one of the twenty possible codons. through 4 are repeated to randomize the third position codon and synthesize the codon at the next position. 5 process is continued until an oligonucleotide of the desired length is achieved. After the final randomization step, the oligonucleotide can be cleaved from the supports and isolated by methods known to one skilled in the art. Alternatively, the oligonucleotides can remain on the supports for use in methods employing probe hybridization.

The diversity of codon sequences, i.e., the number of different possible oligonucleotides, which can be obtained using the methods of the present invention, is extremely large and only limited by the physical characteristics of available materials. For example, a support composed of beads of about 100 μm in diameter will be limited to about 10,000 beads/reaction vessel using a 1 μM reaction vessel containing 25 mg of beads. This size bead can support about 1 x 107 oligonucleotides per bead. Synthesis using separate reaction vessels for each of the twenty amino acids will produce beads in which all the oligonucleotides attached to an individual bead are identical. diversity which can be obtained under these conditions is approximately 107 copies of 10,000 x 20 or 200,000 different random oligonucleotides. The diversity can be increased, however, in several ways without departing from the basic methods disclosed herein. For example, the number of possible sequences can be increased by decreasing the size of the individual beads which make up the support. A bead 30 of about 30 μm in diameter will increase the number of beads per reaction vessel and therefore the number of oligonucleotides synthesized. Another way to increase the diversity of oligonucleotides with random codons is to increase the volume of the reaction vessel. For example, 35 using the same size bead, a larger volume can contain a greater number of beads than a smaller vessel and therefore

number of a greater support the synthesis of oligonucleotides. Increasing the number of codons coupled to a support in a single reaction vessel also increases the diversity of the random oligonucleotides. diversity will be the number of codons coupled per vessel raised to the number of codon positions synthesized. For example, using ten reaction vessels, each synthesizing two codons to randomize a total of twenty codons, the number of different oligonucleotides of ten codons in length per 100 10 μ m bead can be increased where each bead will contain about 2¹⁰ or 1 x 10³ different sequences instead of one. One skilled in the art will know how to modify such parameters to increase the diversity of oligonucleotides with random codons.

A method of synthesizing oligonucleotides having 15 random codons at each position using individual monomers wherein the number of reaction vessels is less than the number of codons to be randomized is also described. example, if twenty codons are to be randomized at each 20 position within an oligonucleotide population, then ten reaction vessels can be used. The use of a smaller number of reaction vessels than the number of codons to be randomized at each position is preferred because the smaller number of reaction vessels is easier to manipulate possible in a greater number of 25 and results oligonucleotides synthesized.

The use of a smaller number of reaction vessels for random synthesis of twenty codons at a desired position within an oligonucleotide is similar to that described above using twenty reaction vessels except that each reaction vessel can contain the synthesis products of more than one codon. For example, st p on synthesis using ten reaction vessels proceeds by coupling about two different codons on supports contained in each of ten reaction vessels. This is shown in Figure 2 where each of the two

relication in

codons coupled to a different support can consist of the following sequences: (1) (T/G)TT for Phe and Val; (2) (T/C)CT for Ser and Pro; (3) (T/C)AT for Tyr and His; (4) (T/C)GT for Cys and Arg; (5) (C/A)TG for Leu and Met; (6) (C/G)AG for Gln and Glu; (7) (A/G)CT for Thr and Ala; (8) (A/G)AT for Asn and Asp; (9) (T/G)GG for Trp and Gly and (10) A(T/A)A for Ile and Cys. The slash (/) signifies that a mixture of the monomers indicated on each side of the slash are used as if they were a single monomer in the indicated coupling step. The antisense sequence for each of the above codons can be generated by synthesizing the complementary sequence. For example, the antisense for Phe and Val can be AA(C/A). The amino acids encoded by each of the above pairs of sequences are given as the standard three letter nomenclature. 15

coupling of the monomers in this fashion will yield codons specifying all twenty of the naturally occurring amino acids attached to supports in ten reaction vessels. However, the number of individual reaction vessels to be used will depend on the number of codons to be randomized at the desired position and can be determined by one skilled in the art. For example, if ten codons are to be randomized, then five reaction vessels can be used for coupling. The codon sequences given above can be used for this synthesis as well. The sequences of the codons can also be changed to incorporate or be replaced by any of the additional forty-four codons which constitutes the genetic code.

The remaining steps of synthesis of oligonucleotides
with random codons using a smaller number of reaction
vessels are as outlined above for synthesis with twenty
reaction v ssels except that the mixing and dividing steps
are performed with supports from about half the number of
reaction vessels. These remaining steps are shown in
figure 2 (steps 2 through 4).

Oligonucleotides having at least one specified tuplet at a predetermined position and the remaining positions . having random tuplets can also be synthesized using the methods described herein. The synthesis steps are similar 5 to those outlined above using twenty or less reaction vessels except that prior to synthesis of the specified codon position, the dividing of the supports into separate reaction vessels for synthesis of different codons is omitted. For example, if the codon at the second position of the oligonucleotide is to be specified, then following synthesis of random codons at the first position and mixing of the supports, the mixed supports are not divided into new reaction vessels but, instead, can be contained in a single reaction vessel to synthesize the specified codon. The specified codon is synthesized sequentially from individual monomers as described above. Thus, the number of reaction vessels can be increased or decreased at each step to allow for the synthesis of a specified codon or a desired number of random codons.

20 Following codon synthesis, the mixed supports are divided into individual reaction vessels for synthesis of the next codon to be randomized (Figure 1, step 3) or can be used without separation for synthesis of a consecutive specified codon. The rounds of synthesis can be repeated for each codon to be added until the desired number of positions with predetermined or randomized codons are obtained.

Synthesis of oligonucleotides with the first position codon being specified can also be synthesized using the above method. In this case, the first position codon is synthesized from the appropriate monomers. The supports ar divid d into the required number of reaction vess ls needed for synthesis of random codons at the second position and the rounds of synthesis, mixing and dividing are performed as described above.

A method of synthesizing oligonucleotides having tuplets which are diverse but biased toward a predetermined sequence is also described herein. This method employs two reaction vessels, one vessel for the synthesis of a predetermined sequence and the second vessel for the synthesis of a random sequence. This method is advantageous to use when a significant number of codon positions, for example, are to be of a specified sequence since it alleviates the use of multiple reaction vessels. Instead, a mixture of four different monomers such as 10 adenine, guanine, cytosine and thymine nucleotides are used for the first and second monomers in the codon. The codon is completed by coupling a mixture of a pair of monomers of either guanine and thymine or cytosine and adenine nucleotides at the third monomer position. In the second vessel, nucleotide monomers are coupled sequentially to yield the predetermined codon sequence. Mixing of the two supports yields a population of oligonucleotides containing both the predetermined codon and the random codons at the Synthesis can proceed by using this desired position. 20 mixture of supports in a single reaction vessel, for example, for coupling additional predetermined codons or, further dividing the mixture into two reaction vessels for synthesis of additional random codons.

The two reaction vessel method can be used for codon synthesis within an oligonucleotide with a predetermined tuplet sequence by dividing the support mixture into two portions at the desired codon position to be randomized. Additionally, this method allows for the extent of randomization to be adjusted. For example, unequal mixing or dividing of the two supports will change the fraction of codons with predetermined sequences compared to those with random codons at the desired position. Unequal mixing and dividing of supports can be useful when there is a need to synthesize random codons at a significant number of positions within an oligonucleotide of a longer or shorter

در (دور) مراوع

 \mathcal{S}_{L}

1

30

length.

The extent of randomization can also be adjusted by using unequal mixtures of monomers in the first, second and third monomer coupling steps of the random codon position. The unequal mixtures can be in any or all of the coupling steps to yield a population of codons enriched in sequences reflective of the monomer proportions.

Synthesis of randomized oligonucleotides is performed using methods well known to one skilled in the art. Linear coupling of monomers can, for example, be accomplished using phosphoramidite chemistry with a MilliGen/Biosearch Cyclone Plus automated synthesizer as described by the manufacturer (Millipore, Burlington, MA). chemistries and automated synthesizers can be employed as 15 well and are known to one skilled in the art.

Synthesis of multiple codons can be performed without modification to the synthesizer by separately synthesizing the codons in individual sets of reactions. Alternatively, modification of an automated DNA synthesizer can be 20 performed for the simultaneous synthesis of codons in multiple reaction vessels.

In one embodiment, the invention provides a plurality of procaryotic cells containing a diverse population of expressible oligonucleotides operationally 25 expression elements, the expressible oligonucleotides having a desirable bias of random codon sequences produced and second diverse combinations first of oligonucleotides having a desirable bias of The invention provides for a method for sequences. constructing such a plurality of procaryotic cells as well.

The oligonucleotides synthesized by the above methods can be used to express a plurality of random peptides which

are unbiased, diverse but biased toward a predetermined sequence or which contain at least one specified codon at a predetermined position. The need will determine which type of oligonucleotide is to be expressed to give the resultant population of random peptides and is known to one skilled in the art. Expression can be performed in any compatible vector/host system. Such systems include, for example, plasmids or phagemids in procaryotes such as E. coli, yeast systems, and other eucaryotic systems such as mammalian cells, but will be described herein in context with its presently preferred embodiment, i.e. expression on Filamentous the surface of filamentous bacteriophage. bacteriophage can be, for example, M13, fl and fd. Such phage have circular single-stranded genomes and double strand replicative DNA forms. Additionally, the peptides can also be expressed in soluble or secreted form depending on the need and the vector/host system employed.

Expression of random peptides on the surface of M13 can be accomplished, for example, using the vector system shown in Figure 3. Construction of the vectors enabling 20 one of ordinary skill to make them are explicitly set out in Examples I and II. The complete nucleotide sequences are given in Figures 5, 6 and 7 (SEQ ID NOS: 1, 2 and 3, produces system This respectively). oligonucleotides functionally linked to expression elements and to gVIII by combining two smaller oligonucleotide portions contained in separate vectors into a single vector. The diversity of oligonucleotide species obtained by this system or others described herein can be 5×10^7 or Diversity of less than 5 x 107 can also be greater. obtained and will be determined by the need and type of random peptides to be expressed. The random combination of two precursor portions into a larger oligonucleotide increases the diversity of the population several fold and has the added advantage of producing oligonucleotides larger than what can be synthesized by standard methods.

Additionally, although the correlation is not known, when the number of possible paths an cligonucleotide can take during synthesis such as described herein is greater than the number of beads, then there will be a correlation between the synthesis path and the sequences obtained. By combining oligonucleotide populations which are synthesized separately, this correlation will be destroyed. Therefore, any bias which may be inherent in the synthesis procedures will be alleviated by joining two precursor portions into a contiguous random oligonucleotide.

Populations of precursor oligonucleotides combined into an expressible form are each cloned into separate vectors. The two precursor portions which make up the combined oligonucleotide corresponds to the carboxy and amino terminal portions of the expressed peptide. Each precursor oligonucleotide can encode either the sense or anti-sense and will depend on the orientation of the expression elements and the gene encoding the fusion portion of the protein as well as the mechanism used to 20 join the two precursor oligonucleotides. For the vectors shown in Figure 3, precursor oligonucleotides corresponding to the carboxy terminal portion of the peptide encode the Those corresponding to the amino terminal sense strand. Oligonucleotide portion encode the anti-sense strand. 25 populations are inserted between the Eco RI and Sac I restriction enzyme sites in M13IX22 and M13IX42 (Figure 3A M13IX42 (SEQ ID NO: 1) is the vector used for sense strand precursor oligonucleotide portions and M13IX22 (SEQ ID NO: 2) is used for anti-sense precursor portions.

The populations of randomized oligonucleotides inserted into the vectors are synthesized with Eco RI and Sac I recognition sequences flanking opposite ends of the random codon sequences. The sits allow annealing and ligation of thes single strand oligonucleotides into a double stranded vector restricted with Eco RI and Sac I.

Alternatively, the oligonucleotides can be inserted into the vector by standard mutagenesis methods. In this latter method, single stranded vector DNA is isolated from the phage and annealed with random oligonucleotides having known sequences complementary to vector sequences. The oligonucleotides are extended with DNA polymerase to produce double stranded vectors containing the randomized oligonucleotides.

The vector used for sense strand oligonucleotice 10 portions, M13IX42 (Figure 3B) contains down-stream and in frame with the Eco RI and Sac I restriction sites a sequence encoding the pseudo-wild type gVIII product. gene encodes the wild type M13 gVIII amino acid sequence but has been changed at the nucleotide level to reduce 15 homologous recombination with the wild type gVIII contained The wild type gVIII is present to on the same vector. ensure that at least some functional, non-fusion coat protein will be produced. The inclusion of a wild type gVIII therefore reduces the possibility of non-viable phage production and biological selection against certain peptide 20 fusion proteins. Differential regulation of the two genes can also be used to control the relative ratio of the pseudo and wild type proteins.

Also contained downstream and in frame with the Eco RI
and Sac I restriction sites is an amber stop codon. The
mutation is located six codons downstream from Sac I and
therefore lies between the inserted oligonucleotides and
the gVIII sequence. As was the function of the wild type
gVIII, the amber stop codon also reduces biological
selection when combining precursor portions to produce
expressible oligonucleotides. This is accomplished by
using a non-suppressor (sup 0) host strain because nonsuppressor strains will terminat expression after the
olig nucleotide sequences but before the pseudo gVIII
sequences. Therefore, the pseudo gVIII will never be

expressed on the phage surface under these circumstances.

Instead, only soluble peptides will be produced.

Expression in a non-suppressor strain can be advantageously utilized when one wishes to produce large populations of soluble peptides. Stop codons other than amber, such as opal and ochre, or molecular switches, such as inducible repressor elements, can also be used to unlink peptide expression from surface expression. Additional controls exist as well and are described below.

The vector used for anti-sense strand oligonucleotide portions, M13IX22, (Figure 3A), contains the expression elements for the peptide fusion proteins. Upstream and in frame with the Sac I and Eco RI sites in this vector is a leader sequence for surface expression. A ribosome binding site and Lac Z promoter/operator elements are present for transcription and translation of the peptide fusion proteins.

Both vectors contain a pair of Fok I restriction enzyme sites (Figure 3 A and B) for joining together two precursor oligonucleotide portions and their vector One site is located at the ends of each 20 sequences. precursor oligonucleotide which is to be joined. second Fok I site within the vectors is located at the end of the vector sequences which are to be joined. The 5' 25 overhang of this second Fok I site has been altered to encode a sequence which is not found in the overhangs produced at the first Fok I site within the oligonucleotide The two sites allow the cleavage of each circular vector into two portions and subsequent ligation of essential components within each vector into a single circular vector where the two oligonucleotide precursor portions form a contiguous sequence (Figure 3C). compatible ov rhangs produced at the two Fok I sites allows conditions to be selected for optimal concatermization or circularization reactions for joining the two vector portions. Such selection of conditions can be used to govern the reaction order and therefore increase the efficiency of joining.

Fok I is a restriction enzyme whose recognition 5 sequence is distal to the point of cleavage. Distal placement of the recognition sequence in its location to the cleavage point is important since if the two were superimposed within the oligonucleotide portions to be combined, it would lead to an invariant codon sequence at the juncture. To alleviate the formation of invariant codons at the juncture, Fok I recognition sequences can be placed outside of the random codon sequence and still be used to restrict within the random sequence. Subsequent annealing of the single-strand overhangs produced by Fok I and ligation of the two oligonucleotide precursor portions 15 allows the juncture to be formed. A variety of restriction enzymes restrict DNA by this mechanism and can be used instead of Fok I to join precursor oligonucleotides without creating invariant codon sequences. Such enzymes include, 20 for example, Alw I, Bbu I, Bsp MI, Hga I, Hph I, Mbo II, Mnl I, Ple I and Sfa NI. One skilled in the art knows how to substitute Fok I recognition sequences for alternative enzyme recognition sequences such as those above, and use precursor joining enzyme for appropriate the oligonucleotide portions. 25

precursor of the sequences Although the invariably have oligonucleotides are random and will oligonucleotides within the two precursor populations whose sequences are sufficiently complementary to anneal after cleavage, the efficiency of annealing can be increased by insuring that the single-strand overhangs within one precursor population will have a complementary sequence This can be within the second precursor population. accomplished by synthesizing a non-degenerate series of known sequences at the Fok I cleavage site coding for each

ti -

. .

: 3

. ...

13

25

of the twenty amino acids. Since the Fok I cleavage site contains a four base overhang, forty different sequences . are needed to randomly encode all twenty amino acids. example, if two precursor populations of ten codons in 5 length are to be combined, then after the ninth codon position is synthesized, the mixed population of supports are divided into forty reaction vessels for each of the populations and complementary sequences for each of the corresponding reaction vessels between populations are The sequences are shown in independently synthesized. Tables III and VI of Example I where the oligonucleotides on columns 1R through 40R form complementary overhangs with the oligonucleotides on the corresponding columns 1L through 40L once cleaved. The degenerate X positions in 15 Table VI are necessary to maintain the reading frame once precursor oligonucleotide portions However, use of restriction enzymes which produce a blunt end, such as Mnl I can be alternatively used in place of Fok I to alleviate the degeneracy introduced in maintaining 20 the reading frame.

The last feature exhibited by each of the vectors is an amber stop codon located in an essential coding sequence within the vector portion lost during combining (Figure 3C). The amber stop codon is present to select for viable phage produced from only the proper combination of precursor oligonucleotides and their vector sequences into a single vector species. Other non-sense mutations or selectable markers can work as well.

The combining step randomly brings together different precursor oligonucleotides within the two populations into a single vector (Figure 3C; M13IX). The vector sequences donated from each independent vector, M13IX22 and M13IX42, are necessary for production of viable phage. Also, since th expression elements are contained in M13IX22 and the 35 gVIII sequences are contained in M13IX42, expression of functional gVIII-peptide fusion proteins cannot be accomplished until the sequences are linked as shown in M13IX.

The combining step is performed by restricting each population of vectors containing randomized oligonucleotides with Fok I, mixing and ligating (Figure 3C). Any vectors generated which contain an amber stop codon will not produce viable phage when introduced into a non-suppressor strain (Figure 3D). Therefore, only the sequences which do not contain an amber stop codon will make up the final population of vectors contained in the library. These vector sequences are the sequences required for surface expression of randomized peptides. By analogous methodology, more than two vector portions can be combined into a single vector which expresses random peptides.

The invention provides for a method of selecting peptides capable of being bound by a ligand binding protein from a population of random peptides by (a) operationally linking a diverse population of first oligonucleotides having a desirable bias of random codon sequences to a first vector; (b) operationally linking population of second oligonucleotides having a desirable bias of random codon sequences to a second vector; (c) combining the vector products of steps (a) and (b) under conditions where said populations of first and second oligonucleotides are joined together into a population of combined vectors; (d) introducing said population of combined vectors into a compatible host under conditions 30 sufficient for expressing said population of random peptides; and (e) determining the peptides which bind to The invention also provides for said binding protein. determining the encoding nucleic acid sequence of such peptides as well.

surface expression of the random peptide library is performed in an amber suppressor strain. As described above, the amber stop codon between the random codon sequence and the gVIII sequence unlinks the two components in a non-suppressor strain. Isolating the phage produced from the non-suppressor strain and infecting a suppressor strain will link the random codon sequences to the gVIII sequence during expression (Figure 3E). Culturing the suppressor strain after infection allows the expression of all peptide species within the library as gVIII-peptide fusion proteins. Alternatively, the DNA can be isolated from the non-suppressor strain and then introduced into a suppressor strain to accomplish the same effect.

The level of expression of gVIII-peptide fusion at controlled additionally be can 15 proteins The gVIII-peptide fusion proteins transcriptional level. inducible control of the are under the promoter/operator system. Other inducible promoters can work as well and are known by one skilled in the art. 20 high levels of surface expression, the suppressor library is cultured in an inducer of the Lac Z promoter such as isopropylthio-B-galactoside (IPTG). Inducible control is beneficial because biological selection against nonfunctional gVIII-peptide fusion proteins can be minimized by culturing the library under non-expressing conditions. Expression can then be induced only at the time of screening to ensure that the entire population of are accurately oligonucleotides within the library represented on the phage surface. Also this can be used to control the valency of the peptide on the phage surface. 30

The surface expression library is screened for specific peptides which bind ligand binding proteins by standard affinity isolation procedures. Such methods include, for exampl, panning, affinity chromatography and solid phase blotting procedures. Panning as described by

parmley and Smith, Gene 73:305-318 (1988), which is incorporated herein by reference, is preferred because high titers of phage can be screened easily, quickly and in small volumes. Furthermore, this procedure can select minor peptide species within the population, which otherwise would have been urdetectable, and amplified to substantially homogenous populations. The selected peptide sequences can be determined by sequencing the nucleic acid encoding such peptides after amplification of the phage population.

The invention provides a plurality of procaryotic cells containing a diverse population of oligonucleotides having a desirable bias of random codon sequences that are operationally linked to expression sequences. The invention provides for methods of constructing such populations of cells as well.

Random oligonucleotides synthesized by any of the methods described previously can also be expressed on the surface of filamentous bacteriophage, such as M13, for example, without the joining together of precursor oligonucleotides. A vector such as that shown in Figure 4, M13IX30, can be used. This vector exhibits all the functional features of the combined vector shown in Figure 3C for surface expression of gVIII-peptide fusion proteins.

The complete nucleotide sequence for M13IX30 (SEQ ID NO: 3) is shown in Figure 7.

M13IX30 contains a wild type gVIII for phage viability and a pseudo gVIII sequence for peptide fusions. The vector also contains in frame restriction sites for cloning random peptides. The cloning sites in this vector are Xho I, Stu I and Spe I. Oligonucleotides should therefore be synthesized with the appropriate complem ntary ends for annealing and ligation or insertional mutagenesis. Alternatively, the appropriate termini can be generated by

pcr technology. Between the restriction sites and the pseudo gVIII sequence is an in-frame amber stop codon, again, ensuring complete viability of phage in constructing and manipulating the library. Expression and screening is performed as described above for the surface expression library of oligonucleotides generated from precursor portions.

Thus, the invention provides a method of selecting peptides capable of being bound by a ligand binding protein from a population of random peptides by (a) operationally linking a diverse population of oligonucleotides having a desirable bias of random codon sequences to expression elements; (b) introducing said population of vectors into a compatible host under conditions sufficient for expressing said population of random peptides; and (c) determining the peptides which bind to said binding protein. Also provided is a method for determining the encoding nucleic acid sequence of such selected peptides.

The following examples are intended to illustrate, but not limit the invention.

EXAMPLE I

Isolation and Characterization of Peptide Ligands Generated From Right and Left Half Random Oligonucleotides

synthesis of random the example shows This 25 oligonucleotides and the construction and expression of surface expression libraries of the encoded randomized peptides. The random peptides of this example derive from joining together of two and mixing Also demonstrated is the isolation and 30 oligonucleotides. characterization of peptide ligands and their corresponding nucl otide sequence for specific binding proteins.

Synthesis of Random Oligonucleotides

The synthesis of two randomized oligonucleotides which correspond to smaller portions of a larger randomized oligonucleotide is shown below. Each of the two smaller 5 portions make up one-half of the larger oligonucleotide. The population of randomized oligonucleotides constituting each half are designated the right and left half. Each population of right and left halves are ten codons in length with twenty random codons at each position. 10 right half corresponds to the sense sequence of the randomized oligonucleotides and encode the carboxy terminal half of the expressed peptides. The left half corresponds sequence of anti-sense oligonucleotides and encode the amino terminal half of the The right and left halves of the expressed peptides. randomized oligonucleotide populations are cloned into separate vector species and then mixed and joined so that the right and left halves come together in random combination to produce a single expression vector species which contains a population of randomized oligonucleotides twenty codons in length. Electroporation of the vector population into an appropriate host produces filamentous phage which express the random peptides on their surface.

The reaction vessels for oligonucleotide synthesis were obtained from the manufacturer of the automated synthesizer (Millipore, Burlington, MilliGen/Biosearch Cyclone Plus Synthesizer). The vessels were supplied as packages containing empty reaction columns (1 μ mole), frits, crimps and plugs (MilliGen/Biosearch Derivatized and underivatized 30 catalog # GEN 860458). control pore glass, phosphoramidite nucleotides, and also were synthesis reagents Crimper and decrimper tools were MilliGen/Biosearch. obtained from Fisher Scientific Co., Pittsburgh, (Catalog numbers 06-406-20 and 06-406-25A, respectively).

Ten reaction columns were used for right half synthesis of random oligonucleotides ten codons in length. The oligonucleotides have 5 monomers at their 3' end of the sequence 5'GAGCT3' and 8 monomers at their 5' end of the sequence 5'AATTCCAT3'. The synthesizer was fitted with a column derivatized with a thymine nucleotide (T-column, MilliGen/Biosearch # 0615.50) and was programmed to synthesize the sequences shown in Table I for each of ten columns in independent reaction sets. The sequence of the last three monomers (from right to left since synthesis proceeds 3' to 5') encode the indicated amino acids:

Table I

		A contract of the contract of	ti di
	Column	Sequence (5' to 3')	Amino Acids
15	column 1R	(T/G)TTGAGCT	Phe and Val
	column 2R	(T/C) CTGAGCT	Ser and Pro
•	column 3R	(T/C) ATGAGCT	Tyr and His
	column 4R	(T/C) GTGAGCT	Cys and Arg
	column 5R	(C/A) TGGAGCT	Leu and Met
20	column 6R	(C/G) AGGAGCT	Gln and Glu
7 -	column 7R	(A/G) CTGAGCT	Thr and Ala
	column 8R	(A/G) ATGAGCT	Asn and Asp
	column 9R	(T/G) GGGAGCT	Trp and Gly
•	column 1R	A(T/A)AGAGCT	Ile and Cys

where the two monomers in parentheses denote a single monomer position within the codon and indicate that an equal mixture of each monomer was added to the reaction for coupling. The monomer coupling reactions for each of the 10 columns were performed as recommended by the manufacturer (amidit version S1.06, # 8400-050990, scale 1 μM). After the last coupling reaction, the columns were washed with acetonitrile and lyophilized to dryness.

Following synthesis, the plugs were removed from each

column using a decrimper and the reaction products were poured into a single weigh boat. Initially the bead mass increases, due to the weight of the monomers, however, at In either later rounds of synthesis material is lost. case, the material was equalized with underivatized control pore glass and mixed thoroughly to obtain a random distribution of all twenty codon species. The reaction products were then aliquotted into 10 new reaction columns by removing 25 mg of material at a time and placing it into separate reaction columns. Alternatively, the reaction products can be aliquotted by suspending the beads in a liquid that is dense enough for the beads to remain dispersed, preferably a liquid that is equal in density to the beads, and then aliquoting equal volumes of the suspension into separate reaction columns. The lip on the inside of the columns where the frits rest was cleared of material using vacuum suction with a syringe and 25 G New frits were placed onto the lips, the plugs were fitted into the columns and were crimped into place 20 using a crimper.

synthesis of the second codon position was achieved using the above 10 columns containing the random mixture of reaction products from the first codon synthesis. The monomer coupling reactions for the second codon position are shown in Table II. An A in the first position means that any monomer can be programmed into the synthesizer. At that position, the first monomer position is not coupled by the synthesizer since the software assumes that the monomer is already attached to the column. An A also denotes that the columns from the previous codon synthesis should be placed on the synthesizer for use in the present synthesis round. Reactions were again sequentially repeated for each column as shown in Table II and the reaction products washed and dried as described above.

Table II

•	Column	Sequence (5' to 3')	Amino Acids
	column 1R	(T/G)TT <u>A</u>	Phe and Val
5	column 2R	(T/C) CT <u>A</u>	Ser and Pro
	column 3R	(T/C)AT <u>A</u>	Tyr and His
	column 4R	(T/C)GT <u>A</u>	Cys and Arg
	column 5R	(C/A)TG <u>A</u>	Leu and Met
	column 6R	(C/G) AG <u>A</u>	Gln and Glu
10	column 7R	(A/G) CT <u>A</u>	Thr and Ala
	column 8R	(A/G) AT <u>A</u>	Asn and Asp
+	column 9R	(T/G)GG <u>A</u>	Trp and Gly
	column 10R	A (T/A) A <u>A</u>	Ile and Cys

Randomization of the second codon position was achieved by removing the reaction products from each of the columns and thoroughly mixing the material. The material was again divided into new reaction columns and prepared for monomer coupling reactions as described above.

Random synthesis of the next seven codons (positions 3 through 9) proceeded identically to the cycle described above for the second codon position and again used the monomer sequences of Table II. Each of the newly repacked columns containing the random mixture of reaction products from synthesis of the previous codon position was used for the synthesis of the subsequent codon position. After synthesis of the codon at position nine and mixing of the reaction products, the material was divided and repacked into 40 different columns and the monomer sequences shown in Table III were coupled to each of the 40 columns in independent reactions. The oligonucleotides from each of the 40 columns were mixed once more and cleav d from the control pore glass as recommended by the manufacturer.

Table III

· ·		
·	Column	Sequence (5' to 3')
	column 1R	AATTCTTTT <u>A</u>
أنتائلين بالسائدية	-column 2R	AATTCTGTTA
5	column 3R	AATTCGTTT <u>A</u>
, in the second	column 4R	AATTCGGTT <u>A</u>
*	column 5R	AATTCTTCT <u>A</u>
	column 6R	AATTCTCCTA
20	column 7R	AATTCGTCT <u>A</u>
10	column 8R	AATTCGCCT <u>A</u>
	column 9R	AATTCTTAT <u>A</u>
	column 10R	AATTCTCATA
	column 11R	AATTCGTAT <u>A</u>
15	column 12R	AATTCGCAT <u>A</u>
	column 13R	AATTCTTGTA
	column 14R	AATTCTCGT <u>A</u>
	column 15R	AATTCGTGT <u>A</u>
v (t	column 16R	AATTCGCGT <u>A</u>
20	column 17R	AATTCTCTG <u>A</u>
	column 18R	AATTCTATG <u>A</u>
	column 19R	AATTCGCTG <u>A</u>
*	column 20R	AATTCGATG <u>A</u>
	column 21R	AATTCTCAGA
25	column 22R	AATTCTGAG <u>A</u>
	column 23R	AATTCGCAG <u>A</u>
	column 24R	AATTCGGAG <u>A</u>
· ·	column 25R	AATTCTACT <u>A</u>
A	column 26R	AATTCTGCTA
30	column 27R	AATTCGACT <u>A</u>
	column 28R	AATTCGGCT <u>A</u>
	column 29R	AATTCTAAT <u>A</u>
	column 30R	AATTCTGAT <u>A</u>
	column 31R	AATTCGAAT <u>A</u>
35	column 32R	
	column 33R	AATTCTTGG <u>A</u>

column	34R	AATTCTGGG <u>A</u>
column	35R	AATTCGTGG <u>A</u>
column	36R	AATTCGGGG <u>A</u>
column	37R	AATTCTATAA
column	38R .	AATTCTAAA <u>A</u>
column	39R	AATTCGATA <u>A</u>
column	40R	AATTCGAAA <u>A</u>

Left half synthesis of random oligonucleotides proceeded similarly to the right half synthesis. This half of the oligonucleotide corresponds to the anti-sense sequence of the encoded randomized peptides. Thus, the complementary sequence of the codons in Tables I through III are synthesized. The left half oligonucleotides also have 5 monomers at their 3' end of the sequence 5'GAGCT3' and 8 monomers at their 5' end of the sequence 5'AATTCCAT3'. The rounds of synthesis, washing, drying, mixing, and dividing are as described above.

fitted with a T-column and programmed to synthesize the sequences shown in Table IV for each of ten columns in independent reaction sets. As with right half synthesis, the sequence of the last three monomers (from right to left) encode the indicated amino acids:

Table IV

•	Sequence	* *
Column	(5' to 3')	Amino Acids
column 1	L AA(A/C)GAGCT	Phe and Val
5 column 2	L AG(A/G)GAGCT	Ser and Pro
column 3		Tyr and His
column 4	L AC(A/G)GAGCT	Cys and Arg
column 5		Leu and Met
column 6		Gln and Glu
10 column 7		Thr and Ala
column 8		Asn and Asp
column 9		Trp and Gly
column 1		Ile and Cys

Following washing and drying, the plugs for each column were removed, mixed and aliquotted into ten new reaction columns as described above. Synthesis of the second codon position was achieved using these ten columns containing the random mixture of reaction products from the first codon synthesis. The monomer coupling reactions for the second codon position are shown in Table V.

Table V

			•	•
		<u>Column</u>	Sequence (5' to 3')	Amino Acids
		column 1L	$AA(A/C)\underline{A}$	Phe and Val
25	*	column 2L	$AG(A/G)\underline{A}$	Ser and Pro
, ;	*	column 3L	AT (A/G) <u>A</u>	Tyr and His
	*	column 4L	AC(A/G)A	Cys and Arg
		column 5L	$CA(G/T)\underline{A}$	Leu and Met
		column 6L	CT(G/C)A	Gln and Glu
30		column 7L	AG (T/C) <u>A</u>	Thr and Ala
30	-	column 8L	$AT(T/C)\underline{A}$	Asn and Asp
		column 9L	CC(A/C) A	Trp and Gly
		column 10L	T(A/T)TA	Ile and Cys

Again, randomization of the second codon position was achieved by removing the reaction products from each of the columns and thoroughly mixing the beads. The beads were repacked into ten new reaction columns.

Random synthesis of the next seven codon positions proceeded identically to the cycle described above for the second codon position and again used the monomer sequences of Table V. After synthesis of the codon at position nine and mixing of the reaction products, the material was divided and repacked into 40 different columns and the monomer sequences shown in Table VI were coupled to each of the 40 columns in independent reactions.

Table VI

15	Column	Sequence (5' to 3')
	column 1L	AATTCCATAAAAXXA
	column 2L	AATTCCATAAACXX <u>A</u>
•	column 3L	AATTCCATAACAXX <u>A</u>
	column 4L	AATTCCATAACCXX <u>A</u>
20	column 5L	AATTCCATAGAAXX <u>A</u>
20	column 6L	AATTCCATAGACXXA
	column 7L	AATTCCATAGGAXXA
	column 8L	AATTCCATAGGCXX <u>A</u>
	column 9L	AATTCCATATAAXX <u>A</u>
25	column 10L	AATTCCATATACXX <u>A</u>
23	column 11L	AATTCCATATGAXX <u>A</u>
	column 12L	AATTCCATATGCXX <u>A</u>
	column 13L	AATTCCATACAAXXA
•	column 14L	AATTCCATACACXX <u>A</u>
30	column 15L	AATTCCATACGAXX <u>A</u>
	column 16L	AATTCCATACGCXX <u>A</u>
	column 17L	AATTCCATCAGAXXA
	column 18L	AATTCCATCAGCXX <u>A</u>
,	column 19L	AATTCCATCATAXXA
35 .	column 20L	AATTCCATCATCXX <u>A</u>

	· .		column	21L	AATTCCATCTGAXX <u>A</u>
	- 1 - ×		column		AATTCCATCTGCXXA
	•		column		AATTCCATCTCAXXA
			column		AATTCCATCTCCXXA
_			column	25L	AATTCCATAGTAXX <u>A</u>
. 5			column		AATTCCATAGTCXXA
	. 59.6 g +21		column		AATTCCATAGCAXXA
,			column		AATTCCATAGCCXXA
	v.		column		AATTCCATATTAXXA
	÷.		column		AATTCCATATTCXXA
10	171		column	31L	AATTCCATATCAXXA
	1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 -		column		AATTCCATATCCXXA
* *			column	33L	AATTCCATCCAAXX <u>A</u>
			column	34L	AATTCCATCCACXXA
	•		column		AATTCCATCCCAXXA
15		~	column		AATTCCATCCCXXA
			column	37L	AATTCCATTATAXXA
			column	38L	AATTCCATTATCXXA
	100		column	<i>:</i> ,	AATTCCATTTTAXXA
20		* * * * * * * * * * * * * * * * * * *	column	40L	AATTCCATTTTCXX <u>A</u>
20				*	=

The first two monomers denoted by an "X" represent an equal mixture of all four nucleotides at that position. This is necessary to retain a relatively unbiased codon sequence at the junction between right and left half oligonucleotides.

The above right and left half random oligonucleotides were cleaved and purified from the supports and used in constructing the surface expression libraries below.

Vector Construction

Two M13-based vectors, M13IX42 (SEQ ID NO: 1) and M13IX22 (SEQ ID NO: 2), were constructed for the cloning and propagation of right and left half populations of random oligonucleotides, respectively. The vectors were specially constructed to facilitate the random joining and subsequent expression of right and left half

oligonucleotide populations. Each vector within the population contains one right and one left half oligonucleotide from the population joined together to form a single contiguous oligonucleotide with random codons which is twenty-two codons in length. The resultant population of vectors are used to construct a surface expression library.

M13IX42, or the right-half vector, was constructed to randomized populations of half right 10 oligonucleotides. M13mp18 (Pharmacia, Piscataway, NJ) was the starting vector. This vector was genetically modified to contain, in addition to the encoded wild type M13 gene VIII already present in the vector: (1) a pseudo-wild type M13 gene VIII sequence with a stop codon (amber) placed 15 between it and an Eco RI-Sac I cloning site for randomized oligonucleotides; (2) a pair of Fok I sites to be used for joining with M13IX22, the left-half vector; (3) a second amber stop codon placed on the opposite side of the vector than the portion being combined with the left-half vector; and (4) various other mutations to remove redundant 20 restriction sites and the amino terminal portion of Lac Z.

The pseudo-wild type M13 gene VIII was used for The pseudo-wild surface expression of random peptides. type gene encodes the identical amino acid sequence as that 25 of the wild type gene; however, the nucleotide sequence has been altered so that only 63% identity exists between this gene and the encoded wild type gene VIII. Modification of the gene VIII nucleotide sequence used for surface homologous possibility of the reduces expression recombination with the wild type gene VIII contained on the same vector. Additionally, the wild type M13 gene VIII was retained in the vector system to ensure that at least some functional, non-fusion coat protein would be produced. The inclusion of wild type gene VIII therefore reduces the possibility of non-viable phage production from the random

peptide fusion genes.

The pseudo-wild type gene VIII was constructed by chemically synthesizing a series of oligonucleotides which encode both strands of the gene. The oligonucleotides are presented in Table VII (SEQ ID NOS: 7 through 16).

TABLE VII

Pseudo-Wild Type Gene VIII Oligonucleotide Series

	Top Strand Oligonucleotides	Sequence (5' to 3')
10	VIII 03	GATCC TAG GCT GAA GGC GAT GAC CCT GCT AAG GCT GC
*	VIII 04	A TTC AAT AGT TTA CAG GCA AGT GCT ACT GAG TAC A
	VIII 05	TT GGC TAC GCT TGG GCT ATG
15	AIII 06	GGT GCT ACC ATA GGG ATT AAA TTA TTC AAA AAG TT
	VIII 07	T ACG AGC AAG GCT TCT TA
20	Bottom Strand Oligonucleotides	*
	VIII 08	AGC TTA AGA AGC CTT GCT CGT
	VIII 09	AAT CCC TAT GGT AGC ACC AAC
25	VIII 10	AGC CCA AGC GTA GCC AAT GTA CTC AGT AGC ACT TG
	VIII 11	C CTG TAA ACT ATT GAA TGC AGC CTT AGC AGG GTC
	VIII 12	ATC GCC TTC AGC CTA G

Except for the terminal oligonucleotides VIII 03 (SEQ

ID NO: 7) and VIII 08 (SEQ ID NO: 12), the above oligonucleotides (oligonucleotides VIII 04-VIII 07 and 09-. 12 (SEQ ID NOS: 8 through 11 and 13 through 16)) were mixed at 200 ng each in 10 μ l final volume and phosphorylated 5 with T4 polynucleotide Kinase (Pharmacia, Piscataway, NJ) with 1 mM ATP at 37°C for 1 hour. The reaction was stopped Terminal oligonucleotides were at 65°C for 5 minutes. added to the mixture and annealed into double-stranded form by heating to 65°C for 5 minutes, followed by cooling to 10 room temperature over a period of 30 minutes. The annealed oligonucleotides were ligated together with 1.0 U of T4 DNA ligase (BRL). The annealed and ligated oligonucleotides yield a double-stranded DNA flanked by a Bam HI site at its 5' end and by a Hind III site at its 3' end. A 15 translational stop codon (amber) immediately follows the Bam HI site. The gene VIII sequence begins with the codon GAA (Glu) two codons 3' to the stop codon. The doublestranded insert was phosphorylated using T4 DNA Kinase (Pharmacia, Piscataway, NJ) and ATP (10 mM Tris-HCl, pH 20 7.5, 10 mM MgCl2) and cloned in frame with the Eco RI and Sac I sites within the M13 polylinker. To do so, M13mp18 was digested with Bam HI (New England Biolabs, Beverley, MA) and Hind III (New England Biolabs) and combined at a molar ratio of 1:10 with the double-stranded insert. 25 ligations were performed at 16°C overnight in 1X ligase buffer (50 mM Tris-HCl, pH 7.8, 10 mM MgCl2, 20 mM DTT, 1 mM ATP, 50 μg/ml BSA) containing 1.0 U of T4 DNA ligase (New England Biolabs). The ligation mixture was transformed into a host and screened for positive clones using standard 30 procedures in the art.

Several mutations were generated within the right-half vector to yield functional M13IX42. The mutations were generated using the method of Kunkel et al., Meth. Enzymol. 154:367-382 (1987), which is incorporated herein by reference, for site-directed mutagenesis. The reagents, strains and protocols were obtained from a Bio Rad

Mutagenesis kit (Bio Rad, Richmond, CA) and mutagenesis was performed as recommended by the manufacturer.

A Fok I site used for joining the right and left halves was generated 8 nucleotides 5' to the unique Eco RI site using the oligonucleotide 5'-CTCGAATTCGTACATCCT The second Fok I site GGTCATAGC-3' (SEQ ID NO: 17). retained in the vector is naturally encoded at position 3547; however, the sequence within the overhang was changed to encode CTTC. Two Fok I sites were removed from the vector at positions 239 and 7244 of M13mp18 as well as the Hind III site at the end of the pseudo gene VIII sequence using the mutant oligonucleotides 5'-CATTTTTGCAGATGGCTTAGA -3' (SEQ ID NO: 18) and 5'-TAGCATTAACGTCCAATA-3' (SEQ ID NO: 19), respectively. New Hind III and Mlu I sites were also introduced at position 3919 and 3951 of M13IX42. 15 oligonucleotides used for this mutagenesis had the sequences 5'-ATATATTTTAGTAAGCTTCATCTTCT-3' (SEQ ID NO: 20) (SEQ ID 5'-GACAAAGAACGCGTGAAAACTTT-3' respectively. The amino terminal portion of Lac Z was deleted by oligonucleotide-directed mutagenesis using the oligonucleotide GCGGGCCTCTTCGCTATTGCTTAAGAAGCCTTGCT-3' (SEQ ID NO: 22). This deletion also removed a third M13mp18 derived Fok I site. The distance between the Eco RI and Sac I sites was increased to ensure complete double digestion by inserting 25 a spacer sequence. The spacer sequence was inserted using oligonucleotide TTCAGCCTAGGATCCGCCGAGCTCTCCTACCTGCGAATTCGTACATCC-3 (SEQID Finally, an amber stop codon was placed at NO: 23). using the mutant oligonucleotide 5'position 4492 TGGATTATACTTCTA AATAATGGA-3' (SEQ ID NO: 24). stop codon is used as a biological selection to ensure the proper recombination of vector sequences to bring together right and left halves of the randomized oligonucleotides. In constructing the above mutations, all changes made in a M13 coding region were performed such that the amino acid

4.

13.

30

sequence remained unaltered. It should be noted that several mutations within M13mp18 were found which differed from the published sequence. Where known, these sequence differences are recorded herein as found and therefore may not correspond exactly to the published sequence of M13mp18.

The sequence of the resultant vector, M13IX42, is shown in Figure 5 (SEQ ID NO: 1). Figure 3A also shows M13IX42 where each of the elements necessary for producing a surface expression library between right and left half randomized oligonucleotides is marked. The sequence between the two Fok I sites shown by the arrow is the portion of M13IX42 which is to be combined with a portion of the left-half vector to produce random oligonucleotides as fusion proteins of gene VIII.

M13IX22, or the left-half vector, was constructed to the left half populations of randomized harbor oligonucleotides. This vector was constructed from M13mp19 (Pharmacia, Piscataway, NJ) and contains: (1) Two Fok I 20 sites for mixing with M13IX42 to bring together the left and right halves of the randomized oligonucleotides; (2) sequences necessary for expression such as a promoter and signal sequence and translation initiation signals; (3) an randomized the for site cloning RI-Sac Eco codon for (4) an amber stop 25 oligonucleotides; and biological selection in bringing together right and left half oligonucleotides.

Of the two Fok I sites used for mixing M13IX22 with M13IX42, one is naturally encoded in M13mp18 and M13mp19 (at position 3547). As with M13IX42, the overhang within this naturally occurring Fok I site was changed to CTTC. The other Fok I site was introduced after construction of the translation initiation signals by site-directed mutagenesis using the oligonucleotide 5'-

TAACACTCATTCCGGATGGAATTCTGGAGTCTGGGT-3' (SEQ ID NO: 25).

The translation initiation signals were constructed by annealing of overlapping oligonucleotides as described above to produce a double-stranded insert containing a 5' Eco RI site and a 3' Hind III site. The overlapping oligonucleotides are shown in Table VIII (SEQ ID NOS: 26 through 34) and were ligated as a double-stranded insert between the Eco RI and Hind III sites of M13mp18 as described for the pseudo gene VIII insert. The ribosome binding site (AGGAGAC) is located in oligonucleotide 015 (SEQ ID NO: 26) and the translation initiation codon (ATG) is the first three nucleotides of oligonucleotide 016 (SEQ ID NO: 27).

TABLE VIII

Oligonucleotide Series for Construction of Translation Signals in M13IX22

Oligonucleotide	Sequence (5' to 3')
015 016 20	AATT C GCC AAG GAG ACA GTC AT AATG AAA TAC CTA TTG CCT ACG GCA GCC GCT GGA TTG TT ATTA CTC GCT GCC CAA CCA GCC ATG
018	GCC GAG CTC GTG AT GACC CAG ACT CCA GATATC CAA CAG
25 019 020	GAA TGA GTG TTA AT TCT AGA ACG CGT C ACGT G ACG CGT TCT AGA AT TAA
021	CACTCA TTC CTG T TG GAT ATC TGG AGT CTG GGT CAT CAC GAG CTC GGC CAT G
30 022	GC TGG TTG GGC AGC GAG TAA TAA
023	GT AGG CAA TAG GTA TTT CAT TAT GAC TGT CCT TGG CG

Oligonucleotide 017 (SEQ ID NO: 27) contained a Sac I restriction site 67 nucleotides downstream from the ATG codon. The naturally occurring Eco RI site was removed and a new site introduced 25 nucleotides downstream from the Sac I. Oligonucleotides 5'-TGACTGTCTCCTTGGCGTGTGAAATTGTTA-3' (SEQ ID NO: 35) and 5'-TAACACTCATTCCGGATGGAATTCTGGAGTCT GGGT-3' (SEQ ID NO: 36) were used to generate each of the mutations, respectively. An amber stop codon was also introduced at position 3263 of M13mp18 using the oligonucleotide 5'-CAATTTTATCCTAAATCTTACCAAC-3' (SEQ ID NO: 37).

In addition to the above mutations, a variety of other modifications were made to remove certain sequences and redundant restriction sites. The LAC Z ribosome binding site was removed when the original Eco RI site in M13mp18 was mutated. Also, the Fok I sites at positions 239, 6361 and 7244 of M13mp18 were likewise removed with mutant oligonucleotides 5'-CATTTTTGCAGATGGCTTAGA-3' (SEQ ID NO: 38), 5'-CGAAAGGGGGGTGTGCTGCAA-3' (SEQ ID NO: 39) and 5'-TAGCATTAACGTCCAATA-3' (SEQ ID NO: 40), respectively. Again, mutations within the coding region did not alter the amino acid sequence.

The resultant vector, M13IX22, is 7320 base pairs in length, the sequence of which is shown in Figure 6 (SEQ ID NO: 2). The Sac I and Eco RI cloning sites are at positions 6290 and 6314, respectively. Figure 3A also shows M13IX22 where each of the elements necessary for producing a surface expression library between right and left half randomized oligonucleotides is marked.

30 Library Construction

Each population of right and left half randomized oligonucleotides from columns 1R through 40R and columns 1L through 40L are cloned separately into M13IX42 and M13IX22,

respectively, to create sublibraries of right and left half randomized oligonucleotides. Therefore, a total of eighty sublibraries are generated. Separately maintaining each population of randomized oligonucleotides until the final screening step is performed to ensure maximum efficiency of annealing of right and left half oligonucleotides. The greater efficiency increases the total number of randomized oligonucleotides which can be obtained. Alternatively, one can combine all forty populations of right half oligonucleotides (columns 1R-40R) into one population and of left half oligonucleotides (columns 1L-40L) into a second population to generate just one sublibrary for each.

For the generation of sublibraries, each of the above populations of randomized oligonucleotides are cloned separately into the appropriate vector. The right half oligonucleotides are cloned into M13IX42 to generate sublibraries M13IX42.1R through M13IX42.40R. The left half oligonucleotides are similarly cloned into M13IX22 to generate sublibraries M13IX22.1L through M13IX22.40L. Each vector contains unique Eco RI and Sac I restriction enzyme sites which produce 5' and 3' single-stranded overhangs, respectively, when digested. The single strand overhangs are used for the annealing and ligation of the complementary single-stranded random oligonucleotides.

The randomized oligonucleotide populations are cloned between the Eco RI and Sac I sites by sequential digestion 25 and ligation steps. Each vector is treated with an excess of Eco RI (New England Biolabs) at 37°C for 2 hours followed by addition of 4-24 units of calf intestinal 30 alkaline phosphatase (Boehringer Mannheim, Indianapolis, IN). Reactions are stopped by phenol/chloroform extraction and ethanol precipitation. The pellets are resuspended in an appropriate amount of distilled or deionized water About 10 pmol of vector is mixed with a 5000-fold (dH₂O). of randomized each population 35 molar excess of

oligonucleotides in 10 μ l of 1% ligase buffer (50 mM Tris-HCl, pH 7.8, 10 mM MgCl₂, 20 mM DTT, 1 mM ATP, 50 μ g/ml BSA) containing 1.0 U of T4 DNA ligase (BRL, Gaithersburg, MD). The ligation is incubated at 16°C for 16 hours. Reactions 5 are stopped by heating at 75°C for 15 minutes and the DNA is digested with an excess of Sac I (New England Biolabs) for 2 hours. Sac I is inactivated by heating at 75°C for 15 minutes and the volume of the reaction mixture is adjusted to 300 μ l with an appropriate amount of 10X ligase 10 buffer and dH20. One unit of T4 DNA ligase (BRL) is added and the mixture is incubated overnight at 16°C. The DNA is ethanol precipitated and resuspended in TE (10 mM Tris-HCl, ligation DNA from each 1 mM EDTA). ,0.8 Hg electroporated into XL1 Blue Cells (Stratagene, La Jolla, CA), as described below, to generate the sublibraries.

E. coli XL1 Blue is electroporated as described by Smith et al., Focus 12:38-40 (1990) which is incorporated herein by reference. The cells are prepared by inoculating a fresh colony of XL1s into 5 mls of SOB without magnesium 20 (20 g bacto-tryptone, 5 g bacto-yeast extract, 0.584 g NaCl, 0.186 g KCl, dH_2O to 1,000 mls) and grown with vigorous aeration overnight at 37°C. SOB without magnesium (500 ml) is inoculated at 1:1000 with the overnight culture and grown with vigorous aeration at 37°C until the OD550 is The cells are harvested by 0.8 (about 2 to 3 h). centrifugation at 5,000 rpm (2,600 x g) in a GS3 rotor (Sorvall, Newtown, CT) at 4°C for 10 minutes, resuspended in 500 ml of ice-cold 10% (v/v) sterile glycerol and centrifuged and resuspended a second time in the same After a third centrifugation, the cells are manner. resuspended in 10% sterile glycerol at a final volume of about 2 ml, such that the OD_{550} of the suspension is 200 to 300. Usually, resuspension is achieved in the 10% glycerolthat remains in the bottle after pouring off the supernate. Cells are frozen in 40 μ l aliquots in microcentrifuge tubes using a dry ice-ethanol bath and stored frozen at -70°C.

Frozen cells are electroporated by thawing slowly on ice before use and mixing with about 10 pg to 500 ng of vector per 40 μ l of cell suspension. A 40 μ l aliquot is placed in an 0.1 cm electroporation chamber (Bio-Rad, Richmond, CA) and pulsed once at 0°C using 200 Ω parallel resistor, 25 μ F, 1.88 kV, which gives a pulse length (τ) of 4 ms. A 10 μ l aliquot of the pulsed cells are diluted into 1 ml SOC (98 mls SOB plus 1 ml of 2 M MgCl₂ and 1 ml of 2 M glucose) in a 12- x 75-mm culture tube, and the culture is shaken at 37°C for 1 hour prior to culturing in selective media, (see below).

methods known to one skilled in the art. Such methods can be found in Sanbrook et al., Molecular Cloning: A Laboratory Manuel, Cold Spring Harbor Laboratory, Cold Spring Harbor, 1989, and in Ausubel et al., Current Protocols in Molecular Biology, John Wiley and Sons, New York, 1989, both of which are incorporated herein by reference. Briefly, the above 1 ml sublibrary cultures were grown up by diluting 50-fold into 2XYT media (16 g tryptone, 10 g yeast extract, 5 g NaCl) and culturing at 37°C for 5-8 hours. The bacteria were pelleted by centrifugation at 10,000 xg. The supernatant containing phage was transferred to a sterile tube and stored at 4°C.

Double strand vector DNA containing right and left half randomized oligonucleotide inserts is isolated from the cell pellet of each sublibrary. Briefly, the pellet is washed in TE (10 mM Tris, pH 8.0, 1 mM EDTA) and recollected by centrifugation at 7,000 rpm for 5' in a sorval centrifuge (Newtown, CT). Pellets are resuspended in 6 mls of 10% Sucrose, 50 mM Tris, pH 8.0. 3.0 ml of 10 mg/µl lysozyne is added and incubated on ice for 20 minutes. 12 mls of 0.2 M NaOH, 1% SDS is added followed by 10 minutes on ice. The suspensions are then incubated on ice for 20 minutes after addition of 7.5 mls of 3 M NaOAC,

pH 4.6. The samples are centrifuged at 15,000 rpm for 15 extracted RNased and at 4°C. minut s phenol/chloroform, followed by ethanol precipitation. The pellets are resuspended, weighed and an equal weight of 5 CsCl, is dissolved into each tube until a density of 1.60° g/ml is achieved. EtBr is added to 600 μ g/ml and the by equilibrium isolated is DNA double-stranded centrifugation in a TV-1665 rotor (Sorval) at 50,000 rpm for 6 hours. These DNAs from each right and left half 10 sublibrary are used to generate forty libraries in which randomized of the halves left and oligonucleotides have been randomly joined together.

Each of the forty libraries are produced by joining together one right half and one left half sublibrary. 15 two sublibraries joined together corresponded to the same half random left right and number for example, sublibrary synthesis. oligonucleotide For M13IX42.1R is joined with M13IX22.1L to produce the surface expression library M13IX.1RL. In the alternative situation where only two sublibraries are generated from the combined 20 populations of all right half synthesis and all left half synthesis, only one surface expression library would be produced.

For the random joining of each right and left half oligonucleotide populations into a single surface expression vector species, the DNAs isolated from each sublibrary are digested an excess of Fok I (New England Biolabs). The reactions are stopped by phenol/chloroform extraction, followed by ethanol precipitation. Pellets are resuspended in dH₂O. Each surface expression library is generated by ligating equal molar amounts (5-10 pmol) of Fok I digested DNA isolated from corresponding right and left half sublibraries in 10 μl of 1X ligase buffer containing 1.0 U of T4 DNA ligase (Bethesda Research Laboratories, Gaithersburg, MD). The ligations proceed

overnight at 16°C and are electroporated into the sup 0 strain MK30-3 (Boehringer Mannheim Biochemical, (BMB), Indianapolis, IN) as previously described for XL1 cells. Because MK30-3 is sup 0, only the vector portions encoding the randomized oligonucleotides which come together will produce viable phage.

Screening of Surface Expression Libraries

Purified phage are prepared from 50 ml liquid cultures of XL1 Blue cells (Stratagene) which are infected at a m.o.i. of 10 from the phage stocks stored at 4°C. cultures are induced with 2 mM IPTG. Supernatants from all cultures are combined and cleared by two centrifugations, and the phage are precipitated by adding 1/7.5 volumes of PEG solution (25% PEG-8000, 2.5 M NaCl), followed by 15 incubation at 4°C overnight. The precipitate is recovered by centrifugation for 90 minutes at 10,000 x g. Phage pellets are resuspended in 25 ml of 0.01 M Tris-HCl, pH 7.6, 1.0 mM EDTA, and 0.1% Sarkosyl and then shaken slowly at room temperature for 30 minutes. The solutions are adjusted to 0.5 M NaCl and to a final concentration of 5% polyethylene glycol. After 2 hours at 4°C, precipitates containing the phage are recovered by centrifugation for 1 hour at 15,000 X g. The precipitates are resuspended in 10 ml of NET buffer (0.1 M NaCl, 1.0 mM 25 EDTA, and 0.01 M Tris-HCl, pH 7.6), mixed well, and the phage repelleted by centrifugation at 170,000 X g for 3 The phage pellets are subsequently resuspended overnight in 2 ml of NET buffer and subjected to cesium chloride centrifugation for 18 hours at 110,000 X g (3.86 g of cesium chloride in 10 ml of buffer). Phage bands are collected, diluted 7-fold with NET buffer, recentrifuged at 170,000 X g for 3 hours, resuspended, and stored at 4°C in 0.3 ml of NET buffer containing 0.1 mM sodium azide.

35

streptavidin coated dishes are first biotinylated and then absorbed against UV-inactivated blocking phage (see below). dissolved reagents are biotinylating dimethylformamide at a ratio of 2.4 mg solid NHS-SS-Biotin 2-(bictinamido) ethyl-1,3'-5 (sulfosuccinimidyl dithiopropionate; Pierce, Rockford, IL) to 1 ml solvent and used as recommended by the manufacturer. Small-scale reactions are accomplished by mixing 1 μ l dissolved reagent with 43 μ l of 1 mg/ml ligand binding protein diluted in sterile bicarbonate buffer (0.1 M NaHCO3, pH 8.6). hours at 25°C, residual biotinylating reagent is reacted with 500 μ l 1 M ethanolamine (pH adjusted to 9 with HCl) for an additional 2 hours. The entire sample is diluted with 1 ml TBS containing 1 mg/ml BSA, concentrated to about 15 50 μ l on a Centricon 30 ultra-filter (Amicon), and washed on the same filter three times with 2 ml TBS and once with 1 ml TBS containing 0.02% NaN3 and 7 x 10^{12} UV-inactivated blocking phage (see below); the final retentate (60-80 μ 1) is stored at 4°C. Ligand binding proteins biotinylated 20 with the NHS-SS-Biotin reagent are linked to biotin via a disulfide-containing chain.

UV-irradiated M13 phage were used for blocking binding proteins which fortuitously bound filamentous phage in M13mp8 (Messing and Vieira, Gene 19: 262-276 general. (1982), which is incorporated herein by reference) was chosen because it carries two amber stop codons, which ensure that the few phage surviving irradiation will not grow in the sup O strains used to titer the surface expression libraries. A 5 ml sample containing 5×10^{13} 30 M13mp8 phage, purified as described above, was placed in a small petri plate and irradiated with a germicidal lamp at a distance of two feet for 7 minutes (flux 150 μ W/cm²). NaN, was added to 0.02% and phage particles concentrated to particl s/ml on a Centricon 30-kDa ultrafilter (Amicon).

For panning, polystyrene petri plates (60 x 15 mm, Falcon; Becton Dickinson, Lincoln Park, NJ) are incubated with 1 ml of 1 mg/ml of streptavidin (BMB) in 0.1 M NaHCO₃ pH 3.6-0.02% NaN₃ in a small, air-tight plastic box overnight in a cold room. The next day streptavidin is removed and replaced with at least 10 ml blocking solution (29 mg/ml of BSA; 3 μg/ml of streptavidin; 0.1 M NaHCO₃ pH 8.6-0.02% NaN₃) and incubated at least 1 hour at room temperature. The blocking solution is removed and plates are washed rapidly three times with Tris buffered saline containing 0.5% Tween 20 (TBS-0.5% Tween 20).

Selection of phage expressing peptides bound by the ligand binding proteins is performed with 5 μ l (2.7 μ g ligand binding protein) of blocked biotinylated ligand binding proteins reacted with a 50 μ l portion of each Each mixture is incubated overnight at 4°C, diluted with 1 ml TBS-0.5% Tween 20, and transferred to a streptavidin-coated petri plate prepared as described above. After rocking 10 minutes at room temperature, 20 unbound phage are removed and plates washed ten times with TBS-0.5% Tween 20 over a period of 30-90 minutes. phage are eluted from plates with 800 μ l sterile elution buffer (1 mg/ml BSA, 0.1 M HCl, pH adjusted to 2.2 with glycerol) for 15 minutes and eluates neutralized with 48 μ l 2 M Tris (pH unadjusted). A 20 μ l portion of each eluate is titered on MK30-3 concentrated cells with dilutions of input phage.

A second round of panning is performed by treating 750 μ l of first eluate from each library with 5 mM DTT for 10 minutes to break disulfide bonds linking biotin groups to residual biotinylated binding proteins. The treated eluate is concentrated on a Centricon 30 ultrafilter (Amicon), washed three times with TBS-0.5% Tween 20, and concentrated to a final volume of about 50 μ l. Final retentate is transferred to a tube containing 5.0 μ l (2.7 μ g ligand)

Å.

Ý

ź

binding protein) blocked biotinylated ligand binding proteins and incubated overnight. The solution is diluted with 1 ml TBS-0.5% Tween 20, panned, and eluted as described above on fresh streptavidin-coated petri plates. The entire second eluate (800 μ l) is neutralized with 48 μ l 2 M Tris, and 20 μ l is titered simultaneously with the first eluate and dilutions of the input phage.

Individual phage populations are purified through 2 to Briefly, the second 3 rounds of plaque purification. eluate titer plates are lifted with nitrocellulose filters (Schleicher & Schuell, Inc., Keene, NH) and processed by washing for 15 minutes in TBS (10 mM Tris-HCl, pH 7.2, 150 mM NaCl), followed by an incubation with shaking for an additional 1 hour at 37°C with TBS containing 5% nonfat dry milk (TBS-5% NDM) at 0.5 ml/cm2. The wash is discarded and fresh TBS-5% NDM is added (0.1 ml/cm2) containing the ligand binding protein between 1 nM to 100 mM, preferably between 1 to 100 μ M. All incubations are carried out in heat-Incubation with the ligand sealable pouches (Sears). binding protein proceeds for 12-16 hours at 4°C with shaking. The filters are removed from the bags and washed 3 times for 30 minutes at room temperature with 150 mls of TBS containing 0.1% NDM and 0.2% NP-40 (Sigma, St. Louis, The filters are then incubated for 2 hours at room temperature in antiserum against the ligand binding protein 25 at an appropriate dilution in TBS-0.5% NDM, washed in 3 changes of TBS containing 0.1% NDM and 0.2% NP-40 as described above and incubated in TBS containing 0.1% NDM and 0.2% NP-40 with 1 x 106 cpm of 125 I-labeled Protein A 30 (specific activity = 2.1 x 10^7 cpm/ μ g). After a washing with TBS containing 0.1% NDM and 0.2% NP-40 as described above, the filters are wrapped in Saran Wrap and exposed to Kodak X-Omat x-ray film (Kodak, Rochester, NY) for 1-12 at -70°C using Dupont Cronex Lightning Plus Intensifying Screens (Dupont, Willmington, DE).

Positive plaques identified are cored with the large end of a pasteur pipet and placed into 1 ml of SM (5.8 g NaCl, 2 g MgSO4.7H2O, 50 ml 1 M Tris-HCl, pH 7.5, 5 mls 2% gelatin, to 1000 mls with dH20) plus 1-3 drops of CHCl3 and incubated at 37°C 2-3 hours or overnight at 4°C. The phage are diluted 1:500 in SM and 2 μ l are added to 300 μ l of XL1 cells plus 3 mls of soft agar per 100 mm2 plate. The XL1 cells are prepared for plating by growing a colony overnight in 10 ml LB (10 g bacto-tryptone, 5 g bacto-yeast extract, 10 g NaCl, 1000 ml dH_20) containing 100 μ l of 20% maltose and 100 μ l of 1 M MgSO $_4$. The bacteria are pelletted by centrifugation at 2000 xg for 10 minutes and the pellet is resuspended gently in 10 mls of 10 mM MgSO4. suspension is diluted 4-fold by adding 30 mls of 10 mM MgSO4 to give an OD_{600} of approximately 0.5. The second and third round screens are identical to that described above except that the plaques are cored with the small end of a pasteur pipet and placed into 0.5 mls SM plus a drop of CHCl3 and 1-5 μl of the phage following incubation are used for plating At the end of the third round of without dilution. purification, an individual plaque is picked and the templates prepared for sequencing.

Template Preparation and Sequencing

Templates are prepared for sequencing by inoculating a 1 ml culture of 2XYT containing a 1:100 dilution of an overnight culture of XL1 with an individual plaque. The plaques are picked using a sterile toothpick. The culture is incubated at 37°C for 5-6 hours with shaking and then transferred to a 1.5 ml microfuge tube. 200 µl of PEG solution is added, followed by vortexing and placed on ice for 10 minutes. The phage precipitate is recovered by centrifugation in a microfuge at 12,000 x g for 5 minutes. The supernatant is discarded and the pellet is resuspended in 230 µl of TE (10 mM Tris-HCl, pH 7.5, 1 mM EDTA) by gently pipeting with a yellow pipet tip. Phenol (200 µl)

is added, followed by a brief vortex and microfuged to separate the phases. The aqueous phase is transferred to extracted with 200 μ l separate tube and phenol/chloroform (1.1) as described above for the phenol 5 extraction. A 0.1 volume of 3 M NaOAc is added, followed by addition of 2.5 volumes of ethanol and precipated at The precipated templates are -20°C for 20 minutes. recovered by centrifugation in a microfuge at 12,000 x g for 8 minutes. The pellet is washed in 70% ethanol, dried 10 and resuspended in 25 μ l TE. Sequencing was performed using a Sequenase sequencing kit following the protocol supplied by the manufacturer (U.S. Biochemical, Cleveland, OH).

EXAMPLE II

15 <u>Isolation and Characterization of Peptide Ligands Generated</u> <u>From Oligonucleotides Having Random Codons at Two</u> <u>Predetermined Positions</u>

This example shows the generation of a surface expression library from a population of oligonucleotides having randomized codons. The oligonucleotides are ten codons in length and are cloned into a single vector species for the generation of a M13 gene VIII-based surface expression library. The example also shows the selection of peptides for a ligand binding protein and characterization of their encoded nucleic acid sequences.

Oligonucleotide Synthesis

30

Oligonucleotides were synthesized as described in Example I. The synthesizer was programmed to synthesize the sequences shown in Table IX. These sequences correspond to the first random codon position synthesized and 3' flanking sequences of the oligonucleotide which hybridizes to the leader sequence in the vector. The

complementary sequences are used for insertional mutagenesis of the synthesized population of oligonucleotides.

Table IX

	Column	Sequence (5' to 3')
5	column 1	AA (A/C) GGTTGGTCGGTACCGG
	column 2	AG(A/G)GGTTGGTCGGTACCGG
	column 3	AT(A/G)GGTTGGTCGGTACCGG
	column 4	AC(A/G)GGTTGGTCGGTACCGG
10	column 5	CA(G/T)GGTTGGTCGGTACCGG
10	column 6	CT(G/C)GGTTGGTCGGTACCGG
	column 7	AG(T/C)GGTTGGTCGGTACCGG
	column 8	AT (T/C) GGTTGGTCGGTACCGG
	column 9	CC(A/C)GGTTGGTCGGTACCGG
15	column 10	T(A/T)TGGTTGGTCGGTACCGG

The next eight random codon positions were synthesized as described for Table V in Example I. Following the ninth position synthesis, the reaction products were once more combined, mixed and redistributed into 10 new reaction columns. Synthesis of the last random codon position and 5' flanking sequences are shown in Table X.

Table X

		Column	Sequence (5' to 3')
		column 1	AGGATCCGCCGAGCTCAA (A/C) A
		column 2	AGGATCCGCCGAGCTCAG (A/G) A
25		column 3	AGGATCCGCCGAGCTCAT(A/G)A
•		column 4	AGGATCCGCCGAGCTCAC(A/G)A
		column 5	AGGATCCGCCGAGCTCCA (G/T) A
•	φ.	column 6	AGGATCCGCCGAGCTCCT(G/C)A
	2-	column 7	AGGATCCGCCGAGCTCAG(T/C)A
30	* * *	column 8	AGGATCCGCCGAGCTCAT(T/C)A
		column 9	AGGATCCGCCGAGCTCCC(A/C)A
•		and the second second	AGGATCCGCCGAGCTCT (A/T) TA
		column 10	AGGATCCGCCGAGCTCT(A/1)1

The reaction products were mixed once more and the oligonucleotides cleaved and purified as recommended by the manufacturer. The purified population of oligonucleotides were used to generate a surface expression library as described below.

Vector Construction

The vector used for generating surface expression libraries from a single oligonucleotide population (i.e., without joining together of right and left half oligonucleotides) is described below. The vector is a M13-based expression vector which directs the synthesis of gene VIII-peptide fusion proteins (Figure 4). This vector exhibits all the functions that the combined right and left half vectors of Example I exhibit.

An M13-based vector was constructed for the cloning 15 populations of surface expression of oligonucleotides (Figure 4, M13IX30), M13mp19 (Pharmacia) This vector was modified to was the starting vector. contain, in addition to the encoded wild type M13 gene 20 VIII: (1) a pseudo-wild type gene, gene VIII sequence with an amber stop codon placed between it and the restriction sites for cloning oligonucleotides; (2) Stu I, Spe I and Xho I restriction sites in frame with the pseudo-wild type gVIII for cloning oligonucleotides; (3) sequences necessary for expression, such as a promoter, signal sequence and translation initiation signals; (4) various other mutations to remove redundant restriction sites and the amino terminal portion of Lac Z.

Construction of M13IX30 was performed in four steps.

30 In the first step, a precursor vector containing the pseudo gene VIII and various other mutations was constructed, M13IX01F. The second step involved th construction of a small cloning site in a separate M13mp18 vector to yield

M13IX03. In the third step, expression sequences and cloning sites were constructed in M13IX03 to generate the intermediate vector M13IX04B. The fourth step involved the incorporation of the newly constructed sequences from the intermediate vector into M13IX01F to yield M13IX30. Incorporation of these sequences linked them with the pseudo gene VIII.

construction of the precursor vector M13IX01F was similar to that of M13IX42 described in Example I except for the following features: (1) M13mp19 was used as the starting vector; (2) the Fok I site 5' to the unique Eco RI site was not incorporated and the overhang at the naturally occurring Fok I site at position 3547 was not changed to 5'-CTTC-3'; (3) the spacer sequence was not incorporated between the Eco RI and Sac I sites; and (4) the amber codon at position 4492 was not incorporated.

In the second step, M13mp18 was mutated to remove the 5' end of Lac Z up to the Lac i binding site and including the Lac Z ribosome binding site and start codon.

20 Additionally, the polylinker was removed and a Mlu I site was introduced in the coding region of Lac Z. A single oligonucleotide was used for these mutagenesis and had the sequence "5'-AAACGACGGCCAGTGCCAAGTGACGCGTGTGAAATTGTTATCC-3'" (SEQ ID NO: 41). Restriction enzyme sites for Hind III and Eco RI were introduced downstream of the MluI site u s i n g the oligonucleotide "5'-GGCGAAAGGGAATTCTGCAAGGCGATTAAGCTTGGGTAACGCC-3'" (SEQ ID NO: 42). These modifications of M13mp18 yielded the vector M13IX03.

30 The expression sequences and cloning sites were introduced into M13IX03 by chemically synthesizing a series of oligonucleotides which encode both strands of the desired sequence. The oligonucleotides are presented in Table XI (SEQ ID NOS: 43 through 50).

TABLE XI
M131X30 Oligonucleotide Series

	*	
	Top Strand Oligonucleotides	Sequence (5' to 3')
5	084	GGCGTTACCCAAGCTTTGTACATGGAGAAAATAAAG
	027	TGAAACAAAGCACTATTGCACTGGCACTCTTACCGT TACCGT
	028	TACTGTTTACCCCTGTGACAAAAGCCGCCCAGGTCC AGCTGC
10	029	TCGAGTCAGGCCTATTGTGCCCAGGGATTGTACTAG TGGATCCG
	Bottom Oligonucleotides	Sequence (5' to 3')
	085	TGGCGAAAGGGAATTCGGATCCACTAGTACAATCCCTG
15	031	GGCACAATAGGCCTGACTCGAGCAGCTGGACCAGGGCG GCTT
	032	TTGTCACAGGGTAAACAGTAACGGTAACGGTAAGTGT GCCA
20	033	GTGCAATAGTGCTTTGTTTCACTTTATTTTCTCCATGT ACAA

The above oligonucleotides except for the terminal oligonucleotides 084 (SEQ ID NO: 43) and 085 (SEQ ID NO: 47) of Table XI were mixed, phosphorylated, annealed and ligated to form a double stranded insert as described in 25 Example I. However, instead of cloning directly into the intermediate vector the insert was first amplified by PCR using the terminal oligonucleotides 084 (SEQ ID NO: 43) and The terminal 085 (SEQ ID NO: 47) as primers. oligonucleotide 084 (SEQ ID NO: 43) contains a Hind III 5 end. its internal to nucleotides 10 30 site Oligonucleotide 085 (SEQ ID NO: 47) has an Eco RI site at its 5' nd. Following amplification, the products were restricted with Hind III and Eco RI and ligated as described in Example I into the polylinker of M13mp18

digested with the same two enzymes. The resultant double stranded insert contained a ribosome binding site, a translation initiation codon followed by a leader sequence and three restriction enzyme sites for cloning random 5 oligonucleotides (Xho I, Stu I, Spe I). The vector was named M13IX04.

During cloning of the double-stranded insert, it was found that one of the GCC codons in oligonucleotides 028 and its complement in 031 was deleted. Since this deletion 10 did not affect function, the final construct is missing one of the two GCC codons. Additionally, oligonucleotide 032 contained a GTG codon where a GAG codon was needed. Mutagenesis was performed using the oligonucleotide 5'-TAACGGTAAGAGTGCCAGTGC-3' (SEQ ID NO: 51) to convert the codon to the desired sequence. The resultant intermediate vector was named M13IX04B.

The fourth step in constructing M13IX30 involved inserting the expression and cloning sequences M13IX04B upstream of the pseudo-wild type gVIII in M13IX01F. This was accomplished by digesting M13IX04B with Dra III and Ban HI and gel isolating the 700 base pair insert containing the sequences of interest. M13IX01F was likewise digested with Dra III and Bam HI. The insert was combined with the double digested vector at a molar ratio of 3:1 and ligated as described in Example I. It should be noted that all modifications in the vectors described herein were confirmed by sequence analysis. The sequence of the final construct, M13IX30, is shown in Figure 7 (SEQ ID NO: 3). Figure 4 also shows M13IX30 where each of the elements necessary for surface expression of randomized oligonucleotides is marked.

Library Construction, Screening and Characterization of Encoded Oligonucleotides

is accomplished identically to that described in Example I for sublibrary construction except the oligonucleotides described above are inserted into M13IX30 by mutagenesis instead of by ligation. The library is constructed and propagated on MK30-3 (BMB) and phage stocks are prepared for infection of XLI cells and screening. The surface expression library is screened and encoding oligonucleotides characterized as described in Example I.

EXAMPLE III

Isolation and Characterization of Peptide Ligands Generated from Right and Left Half Degenerate Oligonucleotides

This example shows the construction and expression of a surface expression library of degenerate oligonucleotides. The encoded peptides of this example derive from the mixing and joining together of two separate oligonucleotide populations. Also demonstrated is the isolation and characterization of peptide ligands and their corresponding nucleotide sequence for specific binding proteins.

Synthesis of Oligonucleotide Populations

A population of left half degenerate oligonucleotides and a population of right half degenerate oligonucleotides was synthesized using standard automated procedures as described in Example I

The d gen rate codon sequences for each population of oligonucleotides were generated by sequentially

synthesizing the triplet NNG/T where N is an equal mixture of all four nucleotides. The antisense sequence for each population of oligonucleotides was synthesized and each population contained 5' and 3' flanking sequences complementary to the vector sequence. The complementary termini was used to incorporate each population of oligonucleotides into their respective vectors by standard mutagenesis procedures. Such procedures have been described previously in Example I and in the Detailed Description. Synthesis of the antisense sequence of each population was necessary since the single-stranded form of the vectors are obtained only as the sense strand.

The left half oligonucleotide population was

synthesized having the following sequence: 5'
AGCTCCCGGATGCCTCAGAAGATG(A/CNN),GGCTTTTGCCACAGGGG-3' (SEQ

ID NO: 52). The right half oligonucleotide population

was synthesized having the following sequence: 5'
CAGCCTCGGATCCGCC(A/CNN),0ATG(A/C)GAAT-3' (SEQ ID NO. 53).

These two oligonucleotide populations when incorporated into their respective vectors and joined together encode a 20 codon oligonucleotide having 19 degenerate positions and an internal predetermined codon sequence.

Vector Construction

Modified forms of the previously described vectors were used for the construction of right and left half sublibraries. The construction of left half sublibraries was performed in an M13-based vector termed M13ED03. This vector is a modified form of the previously described M13IX30 vector and contains all the essential features of both M13IX30 and M13IX22. M13ED03 contains, in addition to a wild type and a pseudo-wild type gene VIII, sequences necessary for expression and two Fok I sites for joining with a right half oligonucleotide

30

sublibrary. Therefore, this vector combines the advantages of both previous vectors in that it can be used for the generation and expression of surface expression libraries from a single oligonucleotide population or it can be joined with a sublibrary to bring together right and left half oligonucleotide populations into a surface expression library.

M13ED03 was constructed in two steps from M13IX30.

The first step involved the modification of M13IX30 to

remove a redundant sequence and to incorporate a sequence encoding the eight amino-terminal residues of human \$\beta\$
endorphin. The leader sequence was also mutated to increase secretion of the product.

During construction of M13IX04 (an intermediate

vector to M13IX30 which is described in Example II), a

six nucleotide sequence was duplicated in oligonucleotide

027 (SEQ ID NO: 44) and its complement 032 (SEQ ID NO:

49). This sequence, 5'-TTACCG-3', was deleted by

mutagenesis in the construction of M13ED01. The

oligonucleotide used for the mutagenesis was 5'
GGTAAACAGTAACGGTAAGAGTGCCAG-3' (SEQ ID NO: 54). The

mutation in the leader sequence was generated using the

oligonucleotide 5'-GGGCTTTTGCCACAGGGGT-3' (SEQ ID NO:

55). This mutagenesis resulted in the A residue at

position 6353 of M13IX30 being changed to a G residue.

The resultant vector was designated M13IX32.

To generate M13ED01, the nucleotide sequence encoding \$\beta\$-endorphin (8 amino acid residues of \$\beta\$-endorphin plus 3 extra amino acid residues) was incorporated after the leader sequence by mutagenesis. The oligonucleotide used had the following sequence: 5'-AGGGTCATCGCCTTCAGCTCCGGATCCCTCAGAAGTCATAAACCCCCCATAGGCTTTTTGCCAC-3' (SEQ ID NO: 56). This mutagenesis also removed some of the downstream sequences through the Spe

I site.

The second step in the construction of M13ED03
involved vector changes which put the β-endorphin
sequence in frame with the downstream pseudo-gene VIII
sequence and incorporated a Fok I site for joining with a
sublibrary of right half oligonucleotides. This vector
was designed to incorporate oligonucleotide populations
by mutagenesis using sequences complementary to those
flanking or overlapping with the encoded β-endorphin
sequence. The absence of β-endorphin expression after
mutagenesis can therefore be used to measure the
mutagenesis frequency. In addition to the above vector
changes, M13ED03 was also modified to contain an amber
codon at position 3262 for biological selection during
joining of right and left half sublibraries.

The mutations were incorporated using standard mutagenesis procedures as described in Example I. The frame shift changes and Fok I site were generated using the oligonucleotide 5'-

- TCGCCTTCAGCTCCCGGATGCCTCAGAAGCATGAACCCCCCATAGGC-3' (SEQ ID NO: 57). The amber codon was generated using the oligonucleotide 5'-CAATTTTATCCTAAATCTTACCAAC-3' (SEQ ID NO: 58). The full sequence of the resultant vector, M13ED03, is provided in Figure 8 (SEQ ID NO: 4).
- The construction of right half oligonucleotide
 sublibraries was performed in a modified form of the
 M13IX42 vector. The new vector, M13IX421, is identical
 to M13IX42 except that the amber codon between the Eco
 RI-SacI cloning site and the pseudo-gene VIII sequence
 was removed. This change ensures that all expression off
 of the Lac Z promoter produces a peptide-gene VIII fusion
 protein. Removal of the amber codon was performed by
 mutagenesis using the following oligonucleotide: 5'GCCTTCAGCCTCGGATCCGCC-3' (SEQ ID NO: 59). The full

sequence of M13IX421 is shown in Figure 9 (SEQ ID NO: 5).

Library Construction, Screening and Characterization of Encoded Oligonucleotides

A sublibrary was constructed for each of the 5 previously described degenerate populations of oligonucleotides. The left half population of oligonucleotides was incorporated into M13ED03 to generate the sublibrary M13ED03.L and the right half population of oligonucleotides was incorporated into 10 M13IX421 to generate the sublibrary M13IX421.R. Each of the oligonucleotide populations were incorporated into their respective vectors using site-directed mutagenesis as described in Example I. Briefly, the nucleotide sequences flanking the degenerate codon sequences were complementary to the vector at the site of incorporation. The populations of nucleotides were hybridized to singlestranded M13ED03 or M13IX421 vectors and extended with T4 DNA polymerase to generate a double-stranded circular vector. Mutant templates were obtained by uridine 20 selection in vivo, as described by Kunkel et al., supra. Each of the vector populations were electroporated into host cells and propagated as described in Example I.

The random joining of right and left half sublibraries into a single surface expression library was accomplished as described in Example I except that prior to digesting each vector population with Fok I they were first digested with an enzyme that cuts in the unwanted portion of each vector. Briefly, M13ED03.L was digested with Bgl II (cuts at 7094) and M13IX421.R was digested with Hind III (cuts at 3919). Each of the digested populations w re further treat d with alkaline phosphatase to ensure that the ends would not religate and then digested with an excess of Fok I. Ligations, electroporation and propagation of the resultant library

was performed as described in Example I.

The surface expression library was screened for ligand binding proteins using a modified panning 5 procedure. Briefly, 1 ml of the library, about 1012 phage particles, was added to 1-5 μg of the ligand binding protein. The ligand binding protein was either an antibody or receptor globulin (Rg) molecule, Aruffo et al., Cell 61:1303-1313 (1990), which is incorporated herein by reference. Phage were incubated shaking with affinity ligand at room temperature for 1 to 3 hours followed by the addition of 200 μl of latex beads (Biosite, San Diego, CA) which were coated with goatantimouse IgG. This mixture was incubated shaking for an 15 additional 1-2 hours at room temperature. Beads were pelleted for 2 minutes by centrifugation in a microfuge and washed with TBS which can contain 0.1% Tween 20. Three additional washes were performed where the last wash did not contain any Tween 20. The bound phage were then eluted with 200 μ l 0.1 M Glycine-HC1, pH 2.2 for 15 minutes and the beads were spun down by centrifugation. The supernatant-containing phage (eluate) was removed and phage exhibiting binding to the ligand binding protein were further enriched by one-to-two more cycles of panning. Typical yields after the first eluate were about 1 \times 10⁶ - 5 \times 10⁶ pfu. The second and third eluate generally yielded about 5 x 10^6 - 2 x 10^7 pfu and 5 x $10^7 - 1 \times 10^{10}$ pfu, respectively.

The second or third eluate was plated at a suitable
density for plaque identification screening and
sequencing of positive clones (i.e., plated at confluency
for rare clones and 200-500 plaques/plate if pure plaques
were needed). Briefly, plaques grown for about 6 hours
at 37°C and were overlaid with nitrocellulose filters
that had been soaked in 2 mM IPTG and then briefly dried.
The filters remained on the plaques overnight at room

1.3

je

1

temperature, removed and placed in blocking solution for 1-2 hours. Following blocking, the filters were incubated in 1 µg/ml ligand binding protein in blocking solution for 1-2 hours at room temperature. Goat antimouse Ig-coupled alkaline phosphatase (Fisher) was added at a 1:1000 dilution and the filters were rapidly washed with 10 mls of TBS or block solution over a glass vacuum filter. Positive plaques were identified after alkaline phosphatase development for detection.

Screening of the degenerate oligonucleotide library 10 with several different ligand binding proteins resulted in the identification of peptide sequences which bound to each of the ligands. For example, screening with an antibody to B-endorphin resulted in the detection of about 30-40 different clones which essentially all had 15 the core amino acid sequence known to interact with the antibody. The sequences flanking the core sequences were different showing that they were independently derived and not duplicates of the same clone. Screening with an antibody known as 57 gave similar results (i.e., a core 20 consensus sequence was identified but the flanking sequences among the clones were different).

EXAMPLE IV

Generation of a Left Half Random Oligonucleotide Library

This example shows the synthesis and construction of a left half random oligonucleotide library.

A population of random oligonucleotides nine codons in length was synthesized as described in Example I except that different sequences at their 5' and 3' ends were synthesized so that they could be easily inserted into the vector by mutagenesis. Also, the mixing and dividing steps for generating random distributions of

reaction products was performed by the alternative method of dispensing equal volumes of bead suspensions. The liquid chosen that was dense enough for the beads to remain dispersed was 100% acetonitrile.

Briefly, each column was prepared for the first coupling reaction by suspending 22 mg (1µmole) of 48 µmol/g capacity beads (Genta, San Diego, CA) in 0.5 mls of 100% acetonitrile. These beads are smaller than those described in Example I and are derivatized with a guanine nucleotide. They also do not have a controlled pore size. The bead suspension was then transferred to an empty reaction column. Suspensions were kept relatively dispersed by gently pipetting the suspension during transfer. Columns were plugged and monomer coupling reactions were performed as shown in Table XII.

Table XII

* · · ·		
		Sequence
*	<u>Column</u>	(5 to 3')
	*	
*	column 1L	AA(A/C)GGCTTTTGCCACAGG
20	column 2L	AG (A/G) GGCTTTTGCCACAGG
	column 3L	AT (A/G) GGCTTTTGCCACAGG
	column 4L	AC(A/G)GGCTTTTGCCACAGG
	column 5L	CA(G/T)GGCTTTTGCCACAGG
	column 6L	CT (G/C) GGCTTTTGCCACAGG
25	column 7L	AG (T/C) GGCTTTTGCCACAGG
23	column 8L	AT (T/C) GGCTTTTGCCACAGG
	column 9L	CC(A/C)GGCTTTTGCCACAGG
	column 10L	T(A/T)TGGCTTTTGCCACAGG
		

After coupling of the last monomer, the columns were unplugged as described previously and their contents were poured into a 1.5 ml microfuge tube. The columns were rinsed with 100% acetonitrile to recover any remaining beads. The volume used for rinsing was determined so

that the final volume of total bead suspension was about 100 µl for each new reaction column that the beads would be aliquoted into. The mixture was vortexed gently to produce a uniformly dispersed suspension and then divided, with constant pipetting of the mixture, into equal volumes. Each mixture of beads was then transferred to an empty reaction column. The empty tubes were washed with a small volume of 100% acetonitrile and also transferred to their respective columns. Random codon positions 2 through 9 were then synthesized as described in Example I where the mixing and dividing steps were performed using a suspension in 100% acetonitrile. The coupling reactions for codon positions 2 through 9 are shown in Table XIII.

15	•	<u>Table XIII</u>
*	<u>Column</u>	Sequence (5' to 3')
•	column 1L	AA(A/C) <u>A</u>
	column 2L	AG (A/G) <u>A</u>
20	column 3L	AT (A/G) <u>A</u>
	column 4L	AC(A/G)A
	column 5L	$CA(G/T)\underline{A}$
٠.	column 6L	$\mathtt{CT}(G/C)\underline{\mathtt{A}}$
	column 7L	$AG(T/C)\underline{A}$
25	column 8L	AT (T/C) <u>A</u>
23	column 9L	CC(A/C) <u>A</u>
	column 10L	T(A/T)TA

After coupling of the last monomer for the ninth codon position, the reaction products were mixed and a portion was transferred to an empty reaction column. Columns were plugged and the following monomer coupling reactions were performed: 5'-CGGATGCCTCAGAAGCCCCXXA-3' (SEQ ID NO: 60). The resulting population of random oligonucl otides was purified and incorporated by

mutagenesis into the left half vector M13ED04.

M13ED04 is a modified version of the M13ED03 vector described in Example III and therefore contains all the features of that vector. The difference between M13ED03 and M13ED04 is that M13ED04 does not contain the five amino acid sequence (Tyr Gly Gly Phe Met) recognized by anti-\(\beta\)-endorphin antibody. This sequence was deleted by mutagenesis using the oligonucleotide 5'-CGGATGCCTCAGAAGGGCTTTTGCCACAGG (SEQ ID NO: 61). The entire nucleotide sequence of this vector is shown in Figure 10 (SEQ ID NO: 6).

Although the invention has been described with reference to the presently preferred embodiment, it should be understood that various modifications can be made without departing from the spirit of the invention. Accordingly, the invention is limited only by the claims.

3.60

420

SEQUENCE LISTING

	•			
	η,		х	•
(1) GENERAL INFORMATION:				*
(i) APPLICANT: Huse, Willi	.am D.	•		•
(ii) TITLE OF INVENTION: SU RANDOMIZED PEPTIDES	RFACE EXPR	ESSION LIB	RARIES OF	
(iii) NUMBER OF SEQUENCES: 6	1		4	
(iv) CORRESPONDENCE ADDRESS (A) ADDRESSEE: Pretty (B) STREET: 444 South (C) CITY: Los Angeles (D) STATE: California (E) COUNTRY: United St (F) ZIP: 90071	, Schroeder Flower Str	r, Bruegger ceet, Suite	nann & Clark 2000	
(v) COMPUTER READABLE FORM: (A) MEDIUM TYPE: Flopp (B) COMPUTER: IBM PC c (C) OPERATING SYSTEM: (D) SOFTWARE: Patentin	y disk compatible PC-DOS/MS-	DOS 1.0, Versi	on #1.25	
(vi) CURRENT APPLICATION DAT (A) APPLICATION NUMBER (B) FILING DATE: (C) CLASSIFICATION:	A: :		. *	*
(viii) ATTORNEY/AGENT INFORMAT (A) NAME: Campbell, Ca (B) REGISTRATION NUMBER (C) REFERENCE/DOCKET N	thryn A R: 31,815	9072	. 7	were we
(ix) TELECOMMUNICATION INFORM (A) TELEPHONE: (619) 5 (B) TELEFAX: (619) 535	35-9001			•
(2) INFORMATION FOR SEQ ID NO:1:			•	
(i) SEQUENCE CHARACTERISTIC (A) LENGTH: 7294 base; (B) TYPE: nucleic acid (C) STRANDEDNESS: both (D) TOPOLOGY: circular	pairs			
(xi) SEQUENCE DESCRIPTION: S	EQ ID NO:1	:		
AATGCTACTA CTATTAGTAG AATTGATGCO			CC AAATGAAAA	AT 60
ATAGCTAAAC AGGTTATTGA CCATTTGCGA	AATGTATCT	A ATGGTCAA	AC TAAATCTA	CT 120
CGTTCGCAGA ATTGGGAATC AACTGTTACA				
GTTGCATATT TAAAACATGT TGAGCTACAC				
TCTGCAAAAA TGACCTCTTA TCAAAAGGAC				
TOTOOLARAT TOTTO				

TTGGAGTTTG CTTCCGGTCT GGTTCGCTTT GAAGCTCGAA TTAAAACGCG ATATTTGAAG

TCTTTCGGGC TTCCTCTTAA TCTTTTTGAT GCAATCCGCT TTGCTTCTGA CTATAATAGT

TOATTCTCCT TTTCTCAACT GTTTAAAGCA	480
CAGGGTAAAG ACCTGATTTT TGATTTATGG TCATTCTCGT TTTCTGAACT GTTTAAAGCA	540
TTTGAGGGGG ATTCAATGAA TATTTATGAC GATTCCGCAG TATTGGACGC TATCCAGTCT	600
AAACATTTTA CTATTACCCC CTCTGGCAAA ACTTCTTTTG CAAAAGCCTC TCGCTATTTT	660
GGTTTTTATC GTCGTCTGGT AAACGAGGGT TATGATAGTG TTGCTCTAC TATGCCTCGT	720
AATTCCTTTT GGCGTTATGT ATCTGCATTA GTTGAATGTG GTATTCCTAA ATCTCAACTG	780
ATGAATCTTT CTACCTGTAA TAATGTTGTT CCGTTAGTTC GTTTTATTAA CGTAGATTTT	840
TCTTCCCAAC GTCCTGACTG GTATAATGAG CCAGTTCTTA AAATCGCATA AGGTAATTCA	900
CAATGATTAA AGTTGAAATT AAACCATCTC AAGCCCAATT TACTACTCGT TCTGGTGTTT	960
CTCGTCAGGG CAAGCCTTAT TCACTGAATG AGCAGCTTTG TTACGTTGAT TTGGGTAATG	1020
AATATCCGGT TCTTGTCAAG ATTACTCTTG ATGAAGGTCA GCCAGCCTAT GCGCCIGGIC	1080
TGTACACCGT TCATCTGTCC TCTTTCAAAG TTGGTCAGTT CGGTTCCCTT ATGATTGACC	1140
GTCTGCGCCT CGTTCCGGCT AAGTAACATG GAGCAGGTCG CGGATTTCGA CACAATTTAT	,
CAGGCGATGA TACAAATCTC CGTTGTACTT TGTTTCGCGC TTGGTATAAT CGCTGGGGGT	1200
CAAAGATGAG TGTTTTAGTG TATTCTTTCG GCTCTTTCGT TTTAGGTTGG TGCCTTCGTA	1320
GTGGCATTAC GTATTTTACC CGTTTAATGG AAACTTCCTC ATGAAAAAGT CTTTAGTCCT	1380
CAAAGCCTCT GTAGCCGTTG CTACCCTCGT TCCGATGCTG TCTTTCGCTG CTGAGGGTGA	
CGATCCCGCA AAAGCGGCCT TTAACTCCCT GCAAGCCTCA GCGACCGAAT ATATCGGTTA	1440
TGCGTGGGCG ATGGTTGTTG TCATTGTCGG CGCAACTATC GGTATCAAGC TGTTTAAGAA	1560
ATTCACCTCG AAAGCAAGCT GATAAACCGA TACAATTAAA GGCTCCTTTT GGAGCCTTTT	1620
TTTTTGGAGA TTTTCAACGT GAAAAATTA TTATTCGCAA TTCCTTTAGT TGTTCCTTTC	
TATTCTCACT CCGCTGAAAC TGTTGAAAGT TGTTTAGCAA AACCCCATAC AGAAAATTCA	1680
TTTACTAACG TCTGGAAAGA CGACAAAACT TTAGATCGTT ACGCTAACTA TGAGGGTTGT	1740
CTGTGGAATG CTACAGGCGT TGTAGTTTGT ACTGGTGACG AAACTCAGTG TTACGGTACA	1800
TGGGTTCCTA TTGGGCTTGC TATCCCTGAA AATGAGGGTG GTGGCTCTGA GGGTGGCGGT	1860 1920
TETGAGGGTG GCGGTTCTGA GGGTGGCGGT ACTAAACCTC CTGAGTACGG TGATACACCT	
ATTCCGGGCT ATACTTATAT CAACCCTCTC GACGGCACTT ATCCGCCTGG TACTGAGCAA	1980
AACCCCCCTA ATCCTAATCC TTCTCTTGAG GAGTCTCAGC CTCTTAATAC TTTCATGTTT	2040
CACAATAATA GGTTCCGAAA TAGGCAGGGG GCATTAACTG TTTATACGGG CACTGTTACT	2100
CAACCCACTG ACCCCGTTAA AACTTATTAC CAGTACACTC CTGTATCATC AAAAGGGAIG	2160
TATGACGCTT ACTGGAACGG TAAATTCAGA GACTGCGCTT TCCATTCTGG CTTTAATGAA	2220
CATCCATTCG TITGTGAATA TCAAGGCCAA TCGTCTGACC TGCCTCAACC TCCTGTCAAT	2280
CCTCCCGGCG GCTCTGGTGG TGGTTCTGGT GGCGGCTCTG AGGGTGGTGG CTCTGAGGGT	2340
GCCGGTTCTG AGGGTGGCGG CTCTGAGGGA GGCGGTTCCG GTGGTGGCTC TGGTTCCGGT	240
ATGAAAGAT GGGAAAGGT AATAAGGGG CTATGACCGA AAATGCCGAT	246

GAAAACGCGC TACAGTCTGA CGCTAAAGGC AAACTTGATT CTGTCGCTAC TGATTACGGT	2520
GCTGCTATCG ATGGTTCAT TGGTGACGTT TCCGGCCTTG CTAATGGTAA TGGTGCTACT	2580
GGTGATTTTG CTGGCTCTAA TTCCCAAATG GCTCAAGTCG GTGACGGTGA TAATTCACCT	2640
TTAATGAATA ATTTCCGTCA ATATTTACCT TCCCTCCCTC AATCGGTTGA ATGTCGCCCT	2700
TTTGTCTTTA GCGCTGGTAA ACCATATGAA TTTTCTATTG ATTGTGACAA AATAAACTTA	2760
TTCCGTGGTG TCTTTGCGTT TCTTTTATAT GTTGCCACCT TTATGTATGT ATTTTCTACG	2820
TTTGCTAACA TACTGCGTAA TAAGGAGTCT TAATCATGCC AGTTCTTTTG GGTATTCCGT	2880
TATTATTGCG TTTCCTCGGT TTCCTTCTGG TAACTTTGTT CGGCTATCTG CTTACTTTTC	2940
TTAAAAAGGG CTTCGGTAAG ATAGCTATTG CTATTTCATT GTTTCTTGCT CTTATTATTG	3000 -
GGCTTAACTC AATTCTTGTG GGTTATCTCT CTGATATTAG CGCTCAATTA CCCTCTGACT	3060
TTGTTCAGGG TGTTCAGTTA ATTCTCCCGT CTAATGCGCT TCCCTGTTTT TATGTTATTC	3120
TCTCTGTAAA GGCTGCTATT TTCATTTTTG ACGTTAAACA AAAAATCGTT TCTTATTTGG	3180
ATTGGGATAA ATAATATGGC TGTTTATTTT GTAACTGGCA AATTAGGCTC TGGAAAGACG	3240
CTCGTTAGCG TTGGTAAGAT TCAGGATAAA ATTGTAGCTG GGTGCAAAAT AGCAACTAAT	3300
CTTGATTTAA GGCTTCAAAA CCTCCCGCAA GTCGGGAGGT TCGCTAAAAC GCCTCGCGTT	3360
CTTAGAATAC CGGATAAGCC TTCTATATCT GATTTGCTTG CTATTGGGCG CGGTAATGAT	3420
TCCTACGATG AAAATAAAAA CGGCTTGCTT GTTCTCGATG AGTGCGGTAC TTGGTTTAAT	3480
ACCCGTTCTT GGAATGATAA GGAAAGACAG CCGATTATTG ATTGGTTTCT ACATGCTCGT	3540
AAATTAGGAT GGGATATTAT CTTCCTTGTT CAGGACTTAT CTATTGTTGA TAAACAGGCG	3600
CGTTCTGCAT TAGCTGAACA TGTTGTTTAT TGTCGTCGTC TGGACAGAAT TACTTTACCT	3660
TITGTCGGTA CTTTATATTC TCTTATTACT GGCTCGAAAA TGCCTCTGCC TAAATTACAT	3720
GTTGGCGTTG TTAAATATGG CGATTCTCAA TTAAGCCCTA CTGTTGAGCG TTGGCTTTAT	3780
ACTGGTAAGA ATTTGTATAA CGCATATGAT ACTAAACAGG CTTTTTCTAG TAATTATGAT	3840
TCCGGTGTTT ATTCTTATTT AACGCCTTAT TTATCACACG GTCGGTATTT CAAACCATTA	3900
AATTTAGGTC AGAAGATGAA GCTTACTAAA ATATATTTGA AAAAGTTTTC ACGCGTTCTT	3960
TGTCTTGCGA TTGGATTTGC ATCAGCATTT ACATATAGTT ATATAACCCA ACCTAAGCCG	4020
GAGGTTAAAA AGGTAGTCTC TCAGACCTAT GATTTTGATA AATTCACTAT TGACTCTTCT	4080
CAGCGTCTTA ATCTAAGCTA TCGCTATGTT TTCAAGGATT CTAAGGGAAA ATTAATTAAT	4140
AGCGACGATT TACAGAAGCA AGGTTATTCA CTCACATATA TTGATTTATG TACTGTTTCC	4200
ATTAAAAAGG TAATTCAAAT GAAATTGTTA AATGTAATTA ATTTTGTTTT CTTGATGTTT	4260
GTTTCATCAT CTTCTTTTGC TCAGGTAATT GAAATGAATA ATTCGCCTCT GCGCGATTTT	4320
GTAACTTGGT ATTCAAAGCA ATCAGGCGAA TCCGTTATTG TTTCTCCCGA TGTAAAAGGT	4380
ACTGTTACTG TATATTCATC TGACGTTAAA CCTGAAAATC TACGCAATTT CTTTATTTCT	4440
GTTTTACGTG CTAATAATTT TGATATGGTT GGTTCAATTC CTTCCATTAT TTAGAAGTAT	4500

AATCCAAACA ATCAGGATTA TATTGATGAA TTGCCATCAT CTGATAATCA GGAATATGAT 4560 GATAATTCCG CTCCTTCTGG TGGTTTCTTT GTTCCGCAAA ATGATAATGT TACTCAAACT 4620 TTTAAAATTA ATAACGTTCG GGCAAAGGAT TTAATACGAG TTGTCGAATT GTTTGTAAAG 4680 TCTAATACTT CTAAATCCTC AAATGTATTA TCTATTGACG GCTCTAATCT ATTAGTTGTT 4740 AGTGCACCTA AAGATATTTT AGATAACCTT CCTCAATTCC TTTCTACTGT TGATTTGCCA ACTGACCAGA TATTGATTGA GGGTTTGATA TTTGAGGTTC AGCAAGGTGA TGCTTTAGAT 4860 TTTTCATTIG CTGCTGGCTC TCAGCGTGGC ACTGTTGCAG GCGGTGTTAA TACTGAGCGC 4920 CTCACCTCTG TTTTATCTTC TGCTGGTGGT TCGTTCGGTA TTTTTAATGG CGATGTTTTA 4980 GGGCTATCAG TTCGCGCATT AAAGACTAAT AGCCATTCAA AAATATTGTC TGTGCCACGT 5040 ATTCTTACGC TTTCAGGTCA GAAGGGTTCT ATCTCTGTTG GCCAGAATGT CCCTTTTATT 5100 ACTGGTCGTG TGACTGGTGA ATCTGCCAAT GTAAATAATC CATTTCAGAC GATTGAGCGT 5160 CAAAATGTAG GTATTTCCAT GAGCGTTTTT CCTGTTGCAA TGGCTGGCGG TAATATTGTT 5220 CTGGATATTA CCAGCAAGGC CGATAGTTTG AGTTCTTCTA CTCAGGCAAG TGATGTTATT 5280 ACTAATCAAA GAAGTATTGC TACAACGGTT AATTIGCGTG ATGGACAGAC TCTTTTACTC 5340 GGTGGCCTCA CTGATTATAA AAACACTTCT CAAGATTCTG GCGTACCGTT CCTGTCTAAA 5400 ATCCCTTTAA TCGGCCTCCT GTTTAGCTCC CGCTCTGATT CCAACGAGGA AAGCACGTTA 5460 TACGTGCTCG TCAAAGCAAC CATAGTACGC GCCCTGTAGC GGCGCATTAA GCGCGGGGG 5520 TGTGGTGGTT ACGCGCAGCG TGACCGCTAC ACTTGCCAGC GCCCTAGCGC CCGCTCCTTT 5580 CGCTTTCTTC CCTTCCTTTC TCGCCACGTT CGCCGGCTTT CCCCGTCAAG CTCTAAATCG 5640 GGGGCTCCCT TTAGGGTTCC GATTTAGTGC TTTACGGCAC CTCGACCCCA AAAAACTTGA 5700 TTTGGGTGAT GGTTCACGTA GTGGGCCATC GCCCTGATAG ACGGTTTTTC GCCCTTTGAC 5760 CTTGGAGTCC ACGTTCTTTA ATAGTGGACT CTTGTTCCAA ACTGGAACAA CACTCAACCC 5820 TATCTCGGGC TATTCTTTTG ATTTATAAGG GATTTTGCCG ATTTCGGAAC CACCATGAAA 5880 CAGGATTITC GCCTGCTGGG GCAAACCAGC GTGGACCGCT TGCTGCAACT CTCTCAGGGC 5940 CAGGCGTGA AGGGCAATCA GCTGTTGCCC GTCTCGCTGG TGAAAAGAAA AACCACCCTG 6000 GCGCCCAATA CGCAAACCGC CTCTCCCCGC GCGTTGGCCG ATTCATTAAT GCAGCTGGCA 6060 CGACAGGTTT CCCGACTGGA AAGCGGGCAG TGAGCGCAAC GCAATTAATG TGAGTTAGCT 6120 CACTCATTAG GCACCCCAGG CTTTACACTT TATGCTTCCG GCTCGTATGT TGTGTGGAAT 6180 TGTGAGCGGA TAACAATTTC ACACAGGAAA CAGCTATGAC CAGGATGTAC GAATTCGCAG 6240 GTAGGAGAGC TCGGCGGATC CTAGGCTGAA GGCGATGACC CTGCTAAGGC TGCATTCAAT 6300 AGTTTACAGG CAAGTGCTAC TGAGTACATT GGCTACGCTT GGGCTATGGT AGTAGTTATA 6360 CTTGGTGCTA CCATAGGGAT TAAATTATTC AAAAAGTTTA CGAGCAAGGC TTCTTAACCA 6420 GCTGGCGTAA TAGCGAAGAG GCCCGCACCG ATCGCCCTTC CCAACAGTTG CGCAGCCTGA 6480 ATGGCGAATG GCGCTTTGCC TGGTTTCCGG CACCAGAAGC GGTGCCGGAA AGCTGGCTGG 6540

AGTGCGATCT TCCTGAGGCC GATACGGTCG TCGTCCCCTC AAACTGGCAG ATGCACGGTT	660
ACGATGCGCC CATCTACACC AACGTAACCT ATCCCATTAC GGTCAATCCG CCGTTTGTTC	666
CCACGGAGAA TCCGACGGGT TGTTACTCGC TCACATTTAA TGTTGATGAA AGCTGGCTAC	672
AGGAAGGCCA GACGCGAATT ATTT TGATG GCGTTCCTAT TGGTTAAAAA ATGAGCTGAT	- 678
TTAACAAAA TTTAACGCGA ATTTTAACAA AATATTAACG TTTACAATTT AAATATTTGC	6840
TTATACAATC TTCCTGTTTT TGGGGCTTTT CTGATTATCA ACCGGGGTAC ATATGATTGA	6900
CATGCTAGTT TTACGATTAC CGTTCATCGA TTCTCTTGTT TGCTCCAGAC TCTCAGGCAA	6960
TGACCTGATA GCCTTTGTAG ATCTCTCAAA AATAGCTACC CTCTCCGGCA TTAATTTATC	7020
AGCTAGAACG GTTGAATATC ATATTGATGG TGATTTGACT GTCTCCGGCC TTTCTCACCC	7080
TTTTGAATCT TTACCTACAC ATTACTCAGG CATTGCATTT AAAATATATG AGGGTTCTAA	7140
AAATTTTTAT CCTTGCGTTG AAATAAAGGC TTCTCCCGCA AAAGTATTAC AGGGTCATAA	7200
IGTTTTTGGT ACAACCGATT TAGCTTTATG CTCTGAGGCT TTATTGCTTA ATTTTGCTAA	7260
TTCTTTGCCT TGCCTGTATG ATTTATTGGA CGTT	7294

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7320 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: both

(D) TOPOLOGY: circular

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

					•	
AATGCTACTA	CTATTAGTAG	AATTGATGCC	ACCTTTTCAG	CTCGCGCCCC	AAATGAAAAT	60
ATAGCTAAAC	AGGTTATTGA	CCATTTGCGA	AATGTATCTA	ATGGTCAAAC	TAAATCTACT	120
CGTTCGCAGA	ATTGGGAATC	AACTGTTACA	TGGAATGAAA	CTTCCAGACA	CCGTACTTTA	180
GTTGCATATT	TAAAACATGT	TGAGCTACAG	CACCAGATTC	AGCAATTAAG	CTCTAAGCCA	240
TCTGCAAAAA	TGACCTCTTA	TCAAAAGGAG	CAATTAAAGG	TACTCTCTAA	TCCTGACCTG	300
TTGGAGTTTG	CTTCCGGTCT	GCTTCGCTTT	GAAGCTCGAA	TTAAAACGCG	ATATTTGAAG	360
TCTTTCGGGC	TTCCTCTTAA	TCTTTTTGAT	GCAATCCGCT	TTGCTTCTGA	CTATAATAGT	420
CAGGGTAAAG	ACCTGATTTT	TGATTTATGG	TCATTCTCGT	TTTCTGAACT	GTTTAAAGCA	480
TTTGAGGGGG	ATTCAATGAA	TATTTATGAÇ	GATTCCGCAG	TATTGGACGC	TATCCAGTCT	540
AAACATTTTA	CTATTACCCC	CTCTGGCAAA	ACTTCTTTTG	CAAAAGCCTC	TCGCTATTTT	600
GGTTTTTATC	GTCGTCTGGT	AAACGAGGGT	TATGATAGTG	TTGCTCTTAC	TATGCCTCGT	660
AATTCCTTTT	GGCGTTATGT	ATCTGCATTA	GTTGAATGTG	GTATTCCTAA	ATCTCAACTG	720
ATGAATCTTT	CTACCTGTAA	TAATGTTGTT	CCGTTAGTTC	GTTTTATTAA	CGTAGATITT	780
		-			AGGTAATTCA	. 840
_					•	

	000
CAATGATTAA AGTTGAAATT AAACCATCTC AAGCCCAATT TACTACTCGT TCTGGTGTTT	900
CTCGTCAGGG CAAGCCTTAT TCACTGAATG AGCAGCTTTG TTACGTTGAT TTGGGTAATG	960
AATATCCGGT TCTTGTCAAG ATTACTCTTG ATGAAGGTCA GCCAGCCTAT GCGCCTGGTC	1020
TGTACACCGT TCATCTGTCC TCTTTCAAAG TTGGTCAGTT CGGTTCCCTT ATGATTGACC	1080
GTGTGCGCCT CGTTCCGGCT AAGTAACATG GAGCAGGTCG CGGATTTCGA CACAATTTAT	- 1140
CAGGCGATGA TACAAATCTC CGTTGTACTT TGTTTCGCGC TTGGTATAAT CGCTGGGGGT	1200
CAAAGATGAG TGTTTTAGTG TATTCTTTCG CCTCTTTCGT TTTAGGTTGG TGCCTTCGTA	1260
GTGGCATTAC GTATTTTACC CGTTTAATGG AAACTTCCTC ATGAAAAAGT CTTTAGTCCT	1320
CAAAGCCTCT GTAGCCGTTG CTACCCTCGT TCCGATGCTG TCTTTCGCTG CTGAGGGTGA	1380
CGATCCCGCA AAAGCGGCCT TTAACTCCCT GCAAGCCTCA GCGACCGAAT ATATCGGTTA	1440
TGCGTGGGCG ATGGTTGTTG TCATTGTCGG CGCAACTATC GGTATCAAGC TGTTTAAGAA	1500
ATTCACCTCG AAAGCAAGCT GATAAACCGA TACAATTAAA GGCTCCTTTT GGAGCCTTTT	1560
TTTTTGGAGA TTTTCAACGT GAAAAAATTA TTATTCGCAA TTCCTTTAGT TGTTCCTTTC	1620
TATTCTCACT CCGCTGAAAC TGTTGAAAGT TGTTTAGCAA AACCCCATAC AGAAAATTCA	1680
TTTACTAACG TCTGGAAAGA CGACAAAACT TTAGATCGTT ACGCTAACTA TGAGGGTTGT	1740
CTGTGGAATG CTACAGGCGT TGTAGTTTGT ACTGGTGACG AAACTCAGTG TTACGGTACA	1800
TGGGTTCCTA TTGGGCTTGC TATCCCTGAA AATGAGGGTG GTGGCTCTGA GGGTGGCGGT	1860
TCTGAGGGTG GCGGTTCTGA GGGTGGCGGT ACTAAACCTC CTGAGTACGG TGATACACCT	1920
ATTCCGGGCT ATACTTATAT CAACCCTCTC GACGGCACTT ATCCGCCTGG TACTGAGCAA	1980
AACCCCGCTA ATCCTAATCC TTCTCTTGAG GAGTCTCAGC CTCTTAATAC TTTCATGTTT	2040
CAGAATAATA GGTTCCGAAA TAGGCAGGGG GCATTAACTG TTTATACGGG CACTGTTACT	2100
CAAGGCACTG ACCCCGTTAA AACTTATTAC CAGTACACTC CTGTATCATC AAAAGCCATG	2160
TATGACGCTT ACTGGAACGG TAAATTCAGA GACTGCGCTT TCCATTCTGG CTTTAATGAA	2220
GATCCATTCG TTTGTGAATA TCAAGGCCAA TCGTCTGACC TGCCTCAACC TCCTGTCAAT	2280
GCTGGCGGCG GCTCTGGTGG TGGTTCTGGT GGCGGCTCTG AGGGTGGTGG CTCTGAGGGT	2340
GGCGGTTCTG AGGGTGGCGG CTCTGAGGGA GGCGGTTCCG GTGGTGGCTC TGGTTCCGGT	2400
GATTTTGATT ATGAAAAGAT GGCAAACGCT AATAAGGGGG CTATGACCGA AAATGCCGAT	2460
GAAAACGCGC TACAGTCTGA CGCTAAAGGC AAACTTGATT CTGTCGCTAC TGATTACGGT	2520
GCTGCTATCG ATGGTTTCAT TGGTGACGTT TCCGGCCTTG CTAATGGTAA TGGTGCTACT	2580
GGTGATTTTG CTGGCTCTAA TTCCCAAATG GCTCAAGTCG GTGACGGTGA TAATTCACCT	2640
TTAATGAATA ATTTCCGTCA ATATTTACCT TCCCTCCCTC AATCGGTTGA ATGTCGCCCT	2700
TTTGTCTTTA GCGCTGGTAA ACCATATGAA TTTTCTATTG ATTGTGACAA AATAAACTTA	2760
TTCCGTGGTG TCTTTGCGTT TCTTTTATAT GTTGCCACCT TTATGTATGT ATTTTCTACG	2820
TITGCTAACA TACTGCGTAA TAAGGAGTCT TAATCATGCC AGTTCTTTTG GGTATTCCGT	2880

ينور

1

TATTATTGCG TTTCCTCGGT TTCCTTCTGG TAACTTTGTT CGGCTATCTG CTTACTTTTC 2940 TTAAAAAGGG CTTCGGTAAG ATAGCTATTG CTATTTCATT GTTTCTTGCT CTTATTATTG 3000 GGCTTAACTC AATTCTTGTG GGTTATCTCT CTGATATTAG CGCTCAATTA CCCTCTGACT 3060 TTGTTCAGGG TGTTCAGTTA ATTCTCCCGT CTAATGCGCT TCCCTGTTTT TATGTTATTC 3120 TCTCTGTAAA GGCTGCTATT TTCATTTTTG ACGTTAAACA AAAAATCGTT TCTTATTTGG 3180 3240 ATTGGGATAA ATAATATGGC TGTTTATTTT GTAACTGGCA AATTAGGCTC TGGAAAGACG CTCGTTAGCG TTGGTAAGAT TTAGGATAAA ATTGTAGCTG GGTGCAAAAT AGCAACTAAT 3300 CTTGATTTAA GGCTTCAAAA CCTCCCGCAA GTCGGGAGGT TCGCTAAAAC GCCTCGCGTT 3360 CTTAGAATAC CGGATAAGCC TTCTATATCT GATTTGCTTG CTATTGGGCG CGGTAATGAT 3420 TCCTACGATG AAAATAAAAA CGGCTTGCTT GTTCTCGATG AGTGCGGTAC TTGGTTTAAT 3480 3540 ACCCGTTCTT GGAATGATAA GGAAAGACAG CCGATTATTG ATTGGTTTCT ACATGCTCGT AAATTAGGAT GGGATATTAT CTTCCTTGTT CAGGACTTAT CTATTGTTGA TAAACAGGCG 3600 CGTTCTGCAT TAGCTGAACA TGTTGTTTAT TGTCGTCGTC TGGACAGAAT TACTTTACCT 3660 TTTGTCGGTA CTTTATATTC TCTTATTACT GGCTCGAAAA TGCCTCTGCC TAAATTACAT 3720 GTTGGCGTTG TTAAATATGG CGATTCTCAA TTAAGCCCTA CTGTTGAGCG TTGGCTTTAT 3780 ACTGGTAAGA ATTTGTATAA CGCATATGAT ACTAAACAGG CTTTTTCTAG TAATTATGAT TCCGGTGTTT ATTCTTATTT AACGCCTTAT TTATCACACG GTCGGTATTT CAAACCATTA 3900 AATTTAGGTC AGAAGATGAA ATTAACTAAA ATATATTTGA AAAAGTTTTC TCGCGTTCTT 3960 4020 TGTCTTGCGA TTGGATTTGC ATCAGCATTT ACATATAGTT ATATAACCCA ACCTAAGCCG GAGGTTAAAA AGGTAGTCTC TCAGACCTAT GATTTTGATA AATTCACTAT TGACTCTTCT 4080 CAGCGTCTTA ATCTAAGCTA TCGCTATGTT TTCAAGGATT CTAAGGGAAA ATTAATTAAT 4140 AGCGACGATT TACAGAAGCA AGGTTATTCA CTCACATATA TTGATTTATG TACTGTTTCC 4200 ATTAAAAAAG GTAATTCAAA TGAAATTGTT AAATGTAATT AATTTTGTTT TCTTGATGTT 4260 TGTTTCATCA TCTTCTTTTG CTCAGGTAAT TGAAATGAAT AATTCGCCTC TGCGCGATTT 4320 TGTAACTTGG TATTCAAAGC AATCAGGCGA ATCCGTTATT GTTTCTCCCG ATGTAAAAGG 4380 TACTGTTACT GTATATTCAT CTGACGTTAA ACCTGAAAAT CTACGCAATT TCTTTATTTC 4440 TCTTTTACGT GCTAATAATT TTGATATGGT TGGTTCAATT CCTTCCATAA TTCAGAAGTA 4500 TAATCCAAAC AATCAGGATT ATATTGATGA ATTGCCATCA TCTGATAATC AGGAATATGA 4560 TGATAATTCC GCTCCTTCTG GTGGTTTCTT TGTTCCGCAA AATGATAATG TTACTCAAAC 4620 TTTTAAAATT AATAACGTTC GGGCAAAGGA TTTAATACGA GTTGTCGAAT TGTTTGTAAA 4680 GTCTAATACT TCTAAATCCT CAAATGTATT ATCTATTGAC GGCTCTAATC TATTAGTTGT 4740 4800 TAGTGCACCT AAAGATATTT TAGATAACCT TCCTCAATTC CTTTCTACTG TTGATTTGCC AACTGACCAG ATATTGATTG AGGGTTTGAT ATTTGAGGTT CAGCAAGGTG ATGCTTTAGA 4860 TTTTTCATTT GCTGCTGGCT CTCAGCGTGG CACTGTTGCA GGCGGTGTTA ATACTGACCG 4920

	4980	
CCTCACCTCT GTTTTATCTT CTGCTGGTGG	TTCGTTCGGT ATTITION CTCTGCCACG 5040	٠
AGGGCTATCA GTTCGCGCAT TAAAGACTAA	TAGCCATTCA ARRATATIGI CIGIGOTIO	
TATTCTTACG CTTTCAGGTC AGAAGGGTTC	TATCTCTGTT GGCCAGAATG TCCCTTTTAT 5100	•
TACTGGTCGT GTGACTGGTG AATCTGCCAA	TGTAAATAAT CCATTTOAGA CGATTGAGA	
TCAAAATGTA GGTATTTCCA TGAGCGTTTT	TCCTGTTGCA AIGGCIGGCG GIAMINITES	:
TCTCCATATT ACCAGCAAGG CCGATAGTTT	GAGTTCTTCT ACTCAGGCAA GTGATGTTAT	
TAGTAATCAA AGAAGTATTG CTACAACGGT	TAATTTGCGT GATGGACAGA CTCTTTTACT	
COCTOCCTC ACTGATTATA AAAACACTTC	TCAAGATTCT GGCGTACCGT TCCTGTGTAA	,
AATCCCTTTA ATCGGCCTCC TGTTTAGCTC	CCGCTCTGAT TCCAACGAGG AAAGCACGII	
ATACCTCCTC GTCAAAGCAA CCATAGTACG	CGCCCTGTAG CGGCGCATTA AGCGCGGGGG	
CTCTCCTGT TACGCCCAGC GTGACCGCTA	CACTTGCCAG CGCCCTAGCG CCCGCTCCT1	
TOGGTTTCTT CCCTTCCTTT CTCGCCACGT	TCGCCGGCTT TCCCCGTCAA GCTCTAAATC 5640	
CCCCCCTCCC TTTAGGGTTC CGATTTAGTG	CTTTACGGCA CCTCGACCCC AAAAAACTTG 5/00	٠,
ATTTGGGTGA TGGTTCACGT AGTGGGCCAT	CGCCCTGATA GACGGTTTTT CGCCCTTTGA 5760	
CCTTCGAGTC CACGTTCTTT AATAGTGGAC	TCTTGTTCCA AACTGGAACA ACACTCAAGC 5820	4
CTATCTCGGG CTATTCTTTT GATTTATAAG	GGATTITGCC GATTICGGAA CCACCATCAA 3880	
ACAGGATTIT CGCCTGCTGG GGCAAACCAG	CGTGGACCGC TTGCTGCAAC TCTCTCAGGG 3940	
CCACCCCCTG AAGGGCAATC AGCTGTTGCC	CGTCTCGCTG GTGAAAAGAA AAACCACCCT 6000	
CCCCCCAAT ACGCAAACCG CCTCTCCCCC	CGCGTTGGCC GATTCATTAA TGCAGCTGGC 8000	
ACCACACGTT TCCCGACTGG AAAGCGGGCA	GTGAGCGCAA CGCAATTAAT GTGAGTTAGC 6120	
TCACTCATTA GGCACCCCAG GCTTTACACT	TTATGCTTCC GGCTCGTATG TTGTGTGGAA 6180	
TTCTCACCCC ATAACAATTI CACACGCCA	A GGAGACAGTC ATAATGAAAT ACCTATTGCC 6240	
TACCCCACCC GCTGGATTGT TATTACTCG	C TGCCCAACCA GCCATGGCCG AGCTCGTGAT 6300	
CACCCACACT CCAGAATTCC ATCCGGAAT	G AGTGTTAATT CTAGAACGCG TAAGCIIGGC	
ACTOCCOCTO GTTTTACAAC GTCGTGACT	G GGAAAACCCT GGCGTTACCC AACTTAATCG 6420	
CCTTCCAGCA CACCCCCCTT TCGCCAGCT	G GCGTAATAGC GAAGAGGCCC GCACCGA1CG 0400	
COCTTCCCAA CACTTGCGCA GCCTGAATC	GG CGAATGGCGC TTTGCCTGGT TTGCGGCACC 0340	
ACAACCCCTC CCGGAAAGCT GGCTGGAG	TG CGATCTTCCT GAGGCCGATA CGGTCGTCGT	
CCCCTCAAAC TGGCAGATGC ACGGTTAC	GA TGCGCCCATC TACACCAACG TAACGTATCC	
CATTACCCTC AATCCGCCGT TTGTTCCC	AC GGAGAATCCG ACGGGTTGTT ACTOGCTCAC	
ATTENDATION CATGAAAGCT GGCTACAG	GA AGGCCAGACG CGAATTATTT TTGATGGCG1	
TAAAAATGA GCTGATTI	AA CAAAAATTA ACGCGAATTT TAACAAAATA	
COTTTA CAATTTAAAT ATTIGCT	TAT ACAATCTTCC TGTTTTTGGG GCTTTTCTGA 0900	
THATCAACCC CCCTACATAT GATTGACA	ATG CTAGTTTTAC GATTACCGTT CATGGATTCT	
TIUTOTTION		

CITGTTTGCT	CCAGACTCTC	AGGCAATGAC	CTGATAGCCT	TTGTAGATCT	CTCAAAAATA	7020
GCTACCCTCT	CCGGCATTAA	TTTATCAGCT	AGAACGGTTG	AATATCATAT	TGATGGTGAT	7080
TTGACTGTCT	CCGGCCTTTC	TCACCCTTTT	GAATCTTTAC	CTACACATTA	CTCAGGCATT	7140
GCATTTAAAA	TATATGAGGG	TTCTAAAAAT	TTTTATCCTT	GCGTTGAAAT	AAAGGCTTCT	7200
ĆCCGCAAAAG	TATTACAGGG	TCATAATGTT	TTTGGTACAA	CCGATTTAGC	TTTATGCTCT	7260
GAGGCTTTAT	TGCTTAATTT	TGCTAATTCT	TTGCCTTGCC	TGTATGATTT	ATTGGACGTT	7320

(2) INFORMATION FOR SEQ ID NO:3:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7445 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: both

(D) TOPOLOGY: circular

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

AATGCTACTA CTATTAGTAG AATTGATGCC ACCTTTTCAG CTCGCGCCCC	AAATGAAAAT	. 60
ATAGCTAAAC AGGTTATTGA CCATTTGCGA AATGTATCTA ATGGTCAAAC	TAAATCTACT	120
CGTTCGCAGA ATTGGGAATC AACTGTTACA TGGAATGAAA CTTCCAGACA	CCGTACTTTA	180
GTTGCATATT TAAAACATGT TGAGCTACAG CACCAGATTC AGCAATTAAC	CTCTAAGCCA	240
TCTGCAAAAA TGACCTCTTA TCAAAAGGAG CAATTAAAGG TACTCTCTAA	TCCTGACCTG	300
TTGGAGTTTG CTTCCGGTCT GGTTCGCTTT GAAGCTCGAA TTAAAACGCC	ATATTTGAAG	360
TCTTTCGGGC TTCCTCTTAA TCTTTTTGAT GCAATCCGCT TTGCTTCTGA	CTATAATAGT	420
CAGGGTAAAG ACCTGATTTT TGATTTATGG TCATTCTCGT TTTCTGAACT	GTTTAAAGCA	480
TTTGAGGGGG ATTCAATGAA TATTTATGAC GATTCCGCAG TATTGGACGC	TATCCAGTCT	540
AAACATTTTA CTATTACCCC CTCTGGCAAA ACTTCTTTTG CAAAAGCCTC	TCGCTATTTT	600
GGTTTTTATC GTCGTCTGGT AAACGAGGGT TATGATAGTG TTGCTCTTAC	TATGCCTCGT	660
AATTCCTTTT GGCGTTATGT ATCTGCATTA GTTGAATGTG GTATTCCTAA	ATCTCAACTG	720
ATGAATCTTT CTACCTGTAA TAATGTTGTT CCGTTAGTTC GTTTTATTAA	A CGTAGATTTT	780
TCTTCCCAAC GTCCTGACTG GTATAATGAG CCAGTTCTTA AAATCGCATA	A AGGTAATTCA	840
CAATGATTAA AGTTGAAATT AAACCATCTC AAGCCCAATT TACTACTCG	T TCTGGTGTTT	900
CTCGTCAGGG CAAGCCTTAT TCACTGAATG AGCAGCTTTG TTACGTTGA	T TTGGGTAATG	960
AATATCCGGT TCTTGTCAAG ATTACTCTTG ATGAAGGTCA GCCAGCCTA		1020
TGTACACCGT TCATCTGTCC TCTTTCAAAG TTGGTCAGTT CGGTTCCCT	T ATGATTGACC	1080
GTCTGCGCCT CGTTCCGGCT AAGTAACATG GAGCAGGTCG CGGATTTCG		1140
CAGGCGATGA TACAAATCTC CGTTGTACTT TGTTTCGCGC TTGGTATAA		1200
CAAAGATGAG TGTTTTAGTG TATTCTTTCG CCTCTTTCGT TTTAGGTTG		1260

	1320
GTGGCATTAC GTATTTTACC CGTTTAATGG AAACTTCCTC ATGAAAAAGT CTTTAGTCCT	1380
CAAAGCCTCT GTAGCCGTTG CTACCCTCGT TCCGATGCTC TCTTTCGCTG CTGAGGGTGA	
CGATCCCGCA AAAGCGGCCT TTAACTCCCT GCAAGCCTCA GCGACCGAAT ATATCGGTTA	1440
TGCGTGGGCG ATGGTTGTTG TCATTGTCGG CGCAACTATC GGTATCAAGC TGTTTAAGAA	1500
ATTCACCTCG AAAGCAAGCT GATAAACCGA TACAATTAAA GGCTCCTTTT GGAGCCTTTT	1560
TTTTTGGAGA TTTTCAACGT GAAAAAATTA TTATTCGCAA TTCCTTTAGT TGTTCCTTTC	1620
TATTCTCACT CCGCTGAAAC TGTTGAAAGT TGTTTAGCAA AACCCCATAC AGAAAATTCA	1680
TTTACTAACG TCTGGAAAGA CGACAAAACT TTAGATCGTT ACGCTAACTA TGAGGGTTGT	1740
CTGTGGAATG CTACAGGCGT TGTAGTTTGT ACTGGTGACG AAACTCAGTG TTACGGTACA	1800
TGGGTTCCTA TTGGGCTTGC TATCCCTGAA AATGAGGGTG GTGGCTCTGA GGGTGGCGGT	1860
TCTGAGGGTG GCGGTTCTGA GGGTGGCGGT ACTAAACCTC CTGAGTACGG TGATACACCT	1920
ATTCCGGGCT ATACTTATAT CAACCCTCTC GACGGCACTT ATCCGCCTGG TACTGAGCAA	1980
AACCCCGCTA ATCCTAATCC TTCTCTTGAG GAGTCTCAGC CTCTTAATAC TTTCATGTTT	2040
CAGAATAATA GGTTCCGAAA TAGGCAGGGG GCATTAACTG TTTATACGGG CACTGTTACT	2100
CAAGGCACTG ACCCCGTTAA AACTTATTAC CAGTACACTC CTGTATCATC AAAAGCCATG	2160
TATGACGCTT ACTGGAACGG TAAATTCAGA GACTGCGCTT TCCATTCTGG CTTTAATGAA	2220
CATCCATTCG TTTGTGAATA TCAAGGCCAA TCGTCTGACC TGCCTCAACC TCCTGTCAAT	2280
GCTGGCGGCG GCTCTGGTGG TGGTTCTGGT GGCGGCTCTG AGGGTGGTGG CTCTGAGGGT	2340
GGGGGTTCTG AGGGTGGCGG CTCTGAGGGA GGCGGTTCCG GTGGTGGCTC TGGTTCCGGT	2400
CATTITGATT ATGAAAAGAT GGCAAACGCT AATAAGGGGG CTATGACCGA AAATGCCGAT	2460
GAAAACGCGC TACAGTCTGA CGCTAAAGGC AAACTTGATT CTGTCGCTAC TGATTACGGT	2520
GCTGCTATCG ATGGTTTCAT TGGTGACGTT TCCGGCCTTG CTAATGGTAA TGGTGCTACT	2580
GGTGATTTTG CTGGCTCTAA TTCCCAAATG GCTCAAGTCG GTGACGGTGA TAATTCACCT	2640
TTAATGAATA ATTTCCGTCA ATATTTACCT TCCCTCCCTC AATCGGTTGA ATGTCGCCCT	2700
TTTGTCTTTA GCGCTGGTAA ACCATATGAA TTTTCTATTG ATTGTGACAA AATAAACTTA	2760
TTCCGTGGTG TCTTTGCGTT TCTTTTATAT GTTGCCACCT TTATGTATGT ATTTTCTACG	2820
TETGCTAACA TACTGCGTAA TAAGGAGTCT TAATCATGCC AGTTCTTTTG GGTATTCCGT	2880
TATTATTGCG TTTCCTCGGT TTCCTTCTGG TAACTTTGTT CGGCTATCTG CTTACTTTTC	
TTAAAAAGGG CTTCGGTAAG ATAGCTATTG CTATTTCATT GTTTCTTGCT CTTATTALIG	3000
GCCTTAACTC AATTCTTGTG GGTTATCTCT CTGATATTAG CGCTCAATTA CCCTCTGACT	3060
TTCTTCAGGG TGTTCAGTTA ATTCTCCCGT CTAATGCGCT TCCCTGTTTT TATGTTATTC	3120
TCTCTGTAAA GGCTGCTATT TTCATTTTTG ACGTTAAACA AAAAATCGTT TCTTATTTGG	3180
ATTGGGATAA ATAATATGGC TGTTTATTTT GTAACTGGCA AATTAGGCTC TGGAAAGACG	3240
CTCGTTAGCG TTGGTAAGAT TCAGGATAAA ATTGTAGCTG GGTGCAAAAT AGCAACTAAT	3300

CONTRACTOR OF CONTRACT TOUCHANAC GCCTCGCGTT	3360
CTTGATTTAA GGCTTCAAAA CCTCCCGCAA GTCGGGAGGT TCGCTAAAAC GCCTCGCGTT	3420
CTTAGAATAC CGGATAAGCC TTCTATATCT GATTTGCTTG CTATTGGGCG CGGTAATGAT	3480
TCCTACGATG AAAATAAAAA CGGCTTGCTT GTTCTCGATG AGTGCGGTAC TTGGTTTAAT	3540
ACCCGTTCTT GGAATGATAA GCAAAGACAG CCGATTATTG ATTGGTTTCT ACATGCTCGT	-
AAATTAGGAT GGGATATTAT TTTTCTTGTT CAGGACTTAT CTATTGTTGA TAAACAGGCG	3600
CGTTCTGCAT TAGCTGAACA TGTTGTTTAT TGTCGTCGTC TGGACAGAAT TACTTTACCT	3660
TTTGTCGGTA CTTTATATTC TCTTATTACT GGCTCGAAAA TGCCTCTGCC TAAATTACAT	3720
GTTGGCGTTG TTAAATATGG CGATTCTCAA TTAAGCCCTA CTGTTGAGCG TTGGCTTTAT	3780
ACTGGTAAGA ATTTGTATAA CGCATATGAT ACTAAACAGG CTTTTTCTAG TAATTATGAT	3840
TCCGGTGTTT ATTCTTATTT AACGCCTTAT TTATCACACG GTCGGTATTT CAAACCATTA	3900
AATTTAGGTC AGAAGATGAA GCTTACTAAA ATATATTTGA AAAAGTTTTC ACGCGTTCTT	3960
TGTCTTGCGA TTGGATTTGC ATCAGCATTT ACATATAGTT ATATAACCCA ACCTAAGCCG	4020
GAGGTTAAAA AGGTAGTCTC TCAGACCTAT GATTTTGATA AATTCACTAT TGACTCTTCT	4080
CAGCGTCTTA ATCTAAGCTA TCGCTATGTT TTCAAGGATT CTAAGGGAAA ATTAATTAAT	4140
AGCGACGATT TACAGAAGCA AGGTTATTCA CTCACATATA TTGATTTATG TACTGTTTCC	4200
ATTAAAAAAG GTAATTCAAA TGAAATTGTT AAATGTAATT AATTTTGTTT TCTTGATGTT	4260
TGTTTCATCA TCTTCTTTTG CTCAGGTAAT TGAAATGAAT AATTCGCCTC TGCGCGATTT	4320
TGTAACTTGG TATTCAAAGC AATCAGGCGA ATCCGTTATT GTTTCTCCCG ATGTAAAAGG	4380
TACTGTTACT GTATATTCAT CTGACGTTAA ACCTGAAAAT CTACGCAATT TCTTTATTTC	4440
TGTTTTACGT GCTAATAATT TTGATATGGT TGGTTCAATT CCTTCCATAA TTCAGAAGTA	4500
TAATCCAAAC AATCAGGATT ATATTGATGA ATTGCCATCA TCTGATAATC AGGAATATGA	4560
TGATAATTCC GCTCCTTCTG GTGGTTTCTT TGTTCCGCAA AATGATAATG TTACTCAAAC	4620
TITTAAAATT AATAACGTTC GGGCAAAGGA TTTAATACGA GTTGTCGAAT TGTTTGTAAA	4680
GTCTAATACT TCTAAATCCT CAAATGTATT ATCTATTGAC GGCTCTAATC TATTAGTTGT	4740
TAGTGCACCT AAAGATATTT TAGATAACCT TCCTCAATTC CTTTCTACTG TTGATTTGCC	4800
AACTGACCAG ATATTGATTG AGGGTTTGAT ATTTGAGGTT CAGCAAGGTG ATGCTTTAGA	4860
TTTTTCATTT GCTGCTGGCT CTCAGCGTGG CACTGTTGCA GGCGGTGTTA ATACTGACCG	4920
CCTCACCTCT GTTTTATCTT CTGCTGGTGG TTCGTTCGCT ATTTTTAATG GCGATGTTTT	4980
AGGGCTATCA GTTCGCGCAT TAAAGACTAA TAGCCATTCA AAAATATTGT CTGTGCCACG	5040
TATTCTTACG CTTTCAGGTC AGAAGGGTTC TATCTCTGTT GGCCAGAATG TCCCTTTTAT	5100
	5160
	-5220
TCAAAATGTA GGTATIICA IGAGGGTIII IGGTGTGTTGT ACTCAGGCAA GTGATGTTAT	5280
TCTGGATATT ACCAGCAAGG CCGATAGTTT GAGTTCTTCT ACTCAGGCAA GTGATGTTAT	5340
TACTAATCAA AGAAGTATTG CTACAACGGT TAATTTGCGT GATGGACAGA CTCTTTTACT	

	<i>**</i>	9		COCCTACCCT	TCCTGTCTAA	5400
	CGGTGGCCTC ACTGATTATA	AAAACACTTC	TCAAGATTCT	GGCGTACCGT	10010102121	5460
	AATCCCTTTA ATCGCCTCC	TGTTTAGCTC	CCGCTCTGAT	TCCAACGAGG	ACCCCCCCCC	5520
	ATACGTGCTC GTGAAAGCAA	CCATAGTACG	CGCCCTGTAG	CGGCGCATTA	AGCGCGGCGG	5580
	GTGTGGTGGT TACGCGCAGC	GTGACCGCTA	CACTTGCCAG	CGCCCTAGCG	00000100==	5640
	TCGCTTTCTT CCCTTCCTTT	CTCGCCACGT	TCGCCGGCTT	TCCCCGTCAA	GUTUTAAATC	5700
-	ASSOCITCE TTTAGGGTTC	CGATTTAGTG	CTTTACGGCA	CCTCGACCGC	AAAAAACIIG	5760
	ATTTGGGTGA TGGTTCACGT	AGTGGGCCAT	CGCCCTGATA	GACGGTTTTT	CGCCCTTTGA	5820
	CONTROL CACGTTCTTT	AATAGTGGAC	TCTTGTTCCA	AACTGGAACA	ACACTGAAGG	5880
	CTATCTCCCC CTATTCTTT	GATTTATAAG	GGATTTTGCC	GATTTCGGAA	CCACCATOAA	5940
	ACACCATTTT CGCCTGCTGG	GGCAAACCAG	CGTGGACCGC	TTGCTGCAAC	TOTOTORGGG	6000
	COACCCCCTC AAGGGCAATC	AGCTGTTGCC	CGTCTCGCTG	GTGAAAAGAA	AAACGACGCI	6060
	CCCCCCAAT ACGCAAACCG	CCTCTCCCCG	CGCGTTGGCC	GATTCATTAA	TGCAGCIGGC	6120
	ACCACACGTT TCCCGACTGG	AAAGCGGGCA	GTGAGCGCAA	CGCAATTAAT	GTGAGTTAGC	6180
	TCACTCATTA GGCACCCCAG	GCTTTACACT	TTATGCTTCC	GGCTCGTATG	TTGTGTGGAA	
	TTCTCACCGG ATAACAATTT	CACACGCGTC	ACTTGGCACT	GGCCGTCGTT	TTACAACGIC	6240 6300
	CTCACTGGGA AAACCCTGGG	GTTACCCAAG	CTTTGTACAT	GGAGAAAATA	AAGTGAAACA	6360
	AACCACTATT GCACTGGCAC	C TCTTACCGTT	ACCGTTACTO	TITACCCCTG	TGACAAAAGC	6420
•	CCCCAGGTC CAGCTGCTC	G AGTCAGGCC1	ATTGTGCCCA	GGGGATTGTA	CTAGTGGATC	6480
	CTACCCTGAA GGCGATGAC	C CTGCTAAGG	TGCATTCAAT	AGTTTACAGO	CAAGIGUIAC	6540
	TCACTACATT GGCTACGCT	T GGGCTATGG	I AGTAGTTATA	GTTGGTGCT	A CCATAGGGAI	6600
	TAAATTATTC AAAAAGTTT	A CGAGCAAGG	C TICTTAAGC	ATAGCGAAGA	A GGCCCGCACC	6660
	GATCGCCCTT CCCAACAGT	T GCGCAGCCT	G AATGGCGAA	r GGCGCTTTG	C CTGGTTTCCG	6720
	CONCORDE CCCTCCCC	A AAGCTGGCT	G GAGTGCGAT	C TTCCTGAGG	C CGATACGGIC	1-00
	CHARGE CALACTEG	A GATGCACGG	T TACGATGCG	C CCATCTAGA	C CAACGIAACC	6780
,	TARGESTES CCCTCAAT	CC GCCGTTTGT	T CCCACGGAC	A ATCCGACGC	G TIGITACICS	6900
	CTCA CATTTA ATGTTGAT	GA AAGCTGGC	TA CAGGAAGG	C AGACGCGA	AL IMITITION	6960
	CCCTTCCTA TTGGTTAA	AA AATGAGCT	GA TTTAACAA	AA ATTTAACG	CG AATTITAACA	7020
	AATATTAAC GTTTACAA	TTAAATATT	TG CTTATACA	AT CTTCCTGT	TT TIGGGGCT11	7080
	TOTAL TATE AACCGGG	TA CATATGAT	TG ACATGCTA	GT TTTACGAT	TA CCGTTCATCG	•
	ATTECT TTGCTCC	AGA CTCTCAGO	CA ATGACCTO	AT AGCCTTTC	TA GATCICIONA	*
	A A THA COTAC COTOTOG	GGC ATTAATT	TAT CAGCTAG	AC GGTTGAA	TAT CATALIGATO	, ,
	TOTOTO	CCC CTTTCTC	ACC CTTTTGA	ATC TTTACCT	ACA CATTACTORG	, 200
	TAAAATA	TAT GAGGGTT	CTA AAAATTT	TTA TCCTTGC	GTT GAAATAAAGG	7320
	CTTCTCCCGC AAAAGTA	ATTA CAGGGTO	ATA ATGTTTT	TGG TACAACC	GAT TTAGCTTTAT	- / 300

GCTCTGAGG	C TITATIGCTI	AATTTTGCTA	ATTCTTTGCC	TTGCCTGTAT	GATTTATTGG	744(
ACGTT	***			(X)		7445

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:

 (A) LENGTH: 7409 base pairs

 (B) TYPE: nucleic acid

 (C) STRANDEDNESS: both

 (D) TOPOLOGY: circular

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

AATGCTACTA CTATTAGTAG AATTGATGCC ACCTTTTCAG CTCGCGCCCC AAATGAAAAT	60
ATAGCTAAAC AGGTTATTGA CCATTTGCGA AATGTATCTA ATGGTCAAAC TAAATCTACT	120
CGTTCGCAGA ATTGGGAATC AACTGTTACA TGGAATGAAA CTTCCAGACA CCGTACTTTA	180
GTTGCATATT TAAAACATGT TGAGCTACAG CACCAGATTC AGCAATTAAG CTCTAAGCCA	240
TCTGCAAAAA TGACCTCTTA TCAAAAGGAG CAATTAAAGG TACTCTCTAA TCCTGACCTG	300
TTGGAGTTTG CTTCCGGTCT GGTTCGCTTT GAAGCTCGAA TTAAAACGCG ATATTTGAAG	360
TCTTTCGGGC TTCCTCTTAA TCTTTTTGAT GCAATCCGCT TTGCTTCTGA CTATAATAGT	420
CAGGGTAAAG ACCTGATTTT TGATTTATGG TCATTCTCGT TTTCTGAACT GTTTAAAGCA	480
TTTGAGGGGG ATTCAATGAA TATTTATGAC GATTCCGCAG TATTGGACGC TATCCAGTCT	540
AAACATTTTA CTATTACCCC CTCTGGCAAA ACTTCTTTTG CAAAAGCCTC TCGCTATTTT	600
GGTTTTTATC GTCGTCTGGT AAACGAGGGT TATGATAGTG TTGCTCTTAC TATGCCTCGT	. 660
AATTCCTTTT GGCGTTATGT ATCTGCATTA GTTGAATGTG GTATTCCTAA ATCTCAACTG	720
ATGAATCTTT CTACCTGTAA TAATGTTGTT CCGTTAGTTC GTTTTATTAA CGTAGATTTT	780
TCTTCCCAAC GTCCTGACTG GTATAATGAG CCAGTTCTTA AAATCGCATA AGGTAATTCA	840
CAATGATTAA AGTTGAAATT AAACCATCTC AAGCCCAATT TACTACTCGT TCTGGTGTTT	900
CTCGTCAGGG CAAGCCTTAT TCACTGAATG AGCAGCTTTG TTACGTTGAT TTGGGTAATG	960
AATATCCGGT TCTTGTCAAG ATTACTCTTG ATGAAGGTCA GCCAGCCTAT GCGCCTGGTC	1020
TGTACACCGT TCATCTGTCC TCTTTCAAAG TTGGTCAGTT CGGTTCCCTT ATGATTGACC	1080
GTCTGCGCCT CGTTCCGGCT AAGTAACATG GAGCAGGTCG CGGATTTCGA CACAATTTAT	1140
CAGGCGATGA TACAAATCTC CGTTGTACTT TGTTTCGCGC TTGGTATAAT CGCTGGGGGT	1200
CAAAGATGAG TGTTTTAGTG TATTCTTTCG CCTCTTTCGT TTTAGGTTGG TGCCTTCGTA	1260
GTGGCATTAC GTATTTTACC CGTTTAATGG AAACTTCCTC ATGAAAAAGT CTTTAGTCCT	1320
CAAAGCCTCT GTAGCCGTTG CTACCCTCGT TCCGATGCTG TCTTTCGCTG CTGAGGGTGA	1380
CGATCCCGCA AAAGCGGCCT TTAACTCCCT GCAAGCCTCA GCGACCGAAT ATATCGGTTA	1440
TGCGTGGGCG ATGGTTGTTG TCATTGTCGG CGCAACTATC GGTATCAAGC TGTTTAAGAA	1500

	121	*	. 02			
	ATTCACCTCG AAAGCAAGCT	CATAAACCGA	TACAATTAAA	GGCTCCTTTT	GGAGCCTTTT	1560
	TO SA CAL TOTAL CAACGT	GAAAAAATTA	TTATTCGCAA	TICCITIAGE	1011	1620
	TARGETT OF THE TARGETT AND THE	TCTTGAAAGT	TGTTTAGCAA	AACCCCATAC	AGILLES	1680
	TATTCTCACT CCGCTGAAAC TTTACTAACG TCTGGAAAGA	CGACAAAACT	TTAGATCGTT	ACCCTAACTA	TGAGGGTTGT	1740
,	cm. cacccct	TCTAGTTTGT	ACTGGTGACG	AAACTOAGTG	11110001111	1800
	TGGGTTCCTA TTGGGCTTGC	TATCCCTGAA	AATGAGGGTG	GTGGCTCTGA	GGGTGGCGGT	1860-
	TCTGAGGGTG GCGGTTCTGA	CCCTCCCCGT	ACTAAACCTC	CTGAGTACGG	TGATACACCT	1920
	ATTCCGGGCT ATACTTATAT	CAACCCTCTC	GACGGCACTT	ATCCCCCTCC	TACTGAGCAA	1980
	AACCCCGCTA ATCCTAATCC	TTCTCTTGAG	GAGTCTCAGC	CTCTTAATAC	TTTCATGTTT	2040
	CAGAATAATA GGTTCCGAAA	TAGGCAGGGG	GCATTAACTG	TTTATACGGG	CACTGTTACT	2100
	ACCCCCTTAA	AACTTATTAC	CAGTACACTC	CTGTATCATC	MAMODOLITA	2160
•	TATGACGCTT ACTGGAACGG	TAAATTCAGA	GACTGCGCTT	TCCATTCTGG	CTTTAATGAA	2220
	GATCCATTCG TTTGTGAATA	TCAAGGCCAA	TCGTCTGACC	TGCCTCAACC	TCCTGTCAAT	2280
	CTCTGGTGG	TGGTTCTGGT	GGCGGCTCTG	AGGGTGGTGG	CICIGAGGGI	2340
	ACCOUNTED ACCOMENGE	CTCTGAGGGA	GGCGGTTCCG	GIGGIGGCIC	IGGIIOGGI	2400
•	ATGAAAAGAT	GGCAAACGCT	AATAAGGGGG	CTATGACCGA	AMAIGGOOM	2460
	TACAGTCTGA	CGCTÁAAGGC	AAACTTGATT	CTGTCGCTAC	TGATTACGGT	2520
	COTTOTATEC ATGGTTTCAT	TGGTGACGTT	TCCGGCCTTC	CTAATGGTAA	IGGIGOTHOL	2580
	CTCCCTCTA	TTCCCAAATC	GCTCAAGTCC	GTGACGGTGA	IMALIONOOL	2640
		ATATTTACCT	TCCCTCCCTC	AATCGGTTGA	' WIGIOGOOA	2700 2760
		A ACCATATGA	A TTTTCTATIO	G ATTGTGACAA	WHINWOITH	
	TOTAL	T TCTTTTATA	I GTTGCCACC	TTATGTATG	[Allifornoo	2820 2880
	TACTCCCTA	A TAAGGAGTC	T TAATCATGC	C AGTIGITION	G GGIKI10001	2940
	TTCCTCGG	T TICCTICTG	G TAACTTIGT	T CGGCTATOL	G GIIROIZZZZ	3000
		G ATACCTATI	G CTATTTCAT	T CTTTCTTGC	I CITALITATE	3060
	- ACTO AATTOTTO	G GGTTATCTC	CT CTGATATTA	C CCCLCWWIT	W 000101111	*:-
	TOTTCAGE	TA ATTCTCCC	GT CTAATGCG	CT TCCCTGTT1	II IMIGITALIO	3180
		TT TTCATTIT	TG ACGTTAAA	CA AAAAATUG.	II IOIIRIII	983
		CC TGTTTATT	TT GTAACTGG	CA AATTAGGU	IC. IGGMuteria	
	TTCCTAAG	AT TCAGGATA	AA ATTGTAGC	TG GGTGCAAA	AI ACCIDIO	
	COUTTCAL	AA CCTCCCG	AA GTCGGGAG	GT TCGCTAAA	AC GCCTCGCCTT	36
	THE COCATAN	CC TTCTATAT	CT GATTIGCT	ITG CTATIGGC	,CG	
	TO ALANDA	AAA CGGCTTG	CTT GTTCTCG	ATG AGTGCGG	IAC IIGGIIIII	* * * * * * * * * * * * * * * * * * * *
	ACCOGTTCTT GGAATGA	TAA GGAAAGA	CAG CCGATTA	TTG ATTGGTT	ICL WONIGOTOGI	
					· · · · · · · · · · · · · · · · · · ·	

	AAATTAGGAT GGGATATTAT TTTTCTTGTT CAGGACTTAT CTATTGTTGA TAAACAGGCG	3600
	CGTTCTGCAT TAGCTGAACA TGTTGTTTAT TGTCGTCGTC TGGACAGAAT TACTTTACCT	3660
	TTTGTCGGTA CTTTATATTC TCTTATTACT GGCTCGAAAA TGCCTCTGCC TAAATTACAT	3720
	GTTGGCGTTG TTAAATATGG CGATTCTCAA TTAAGCCCTA CTGTTGAGCG TTGGCTTTAT	3780
	ACTGGTAAGA ATTTGTATAA CGCATATGAT ACTAAACAGG CTTTTTCTAG TAATTATGAT	3840
	TCCGGTGTTT ATTCTTATTT AACGCCTTAT TTATCACACG GTCGGTATTT CAAACCATTA	3900
	AATTTAGGTC AGAAGATGAA GCTTACTAAA ATATATTTGA AAAAGTTTTC ACGCGTTCTT	3960
	TGTCTTGCGA TTGGATTTGC ATCAGCATTT ACATATAGTT ATATAACCCA ACCTAAGCCG	4020
	GAGGTTAAAA AGGTAGTCTC TCAGACCTAT GATTTTGATA AATTCACTAT TGACTCTTCT	4080
	CAGCGTCTTA ATCTAAGCTA TCGCTATGTT TTCAAGGATT CTAAGGGAAA ATTAATTAAT	4140
	AGCGACGATT TACAGAAGCA AGGTTATTCA CTCACATATA TTGATTTATG TACTGTTTCC	4200
	ATTAAAAAAG GTAATTCAAA TGAAATTGTT AAATGTAATT AATTTTGTTT TCTTGATGTT	4260
	TGTTTCATCA TCTTCTTTTG CTCAGGTAAT TGAAATGAAT AATTCGCCTC TGCGCGATTT	4320
	TGTAACTIGG TATTCAAAGC AATCAGGCGA ATCCGTTATT GTTTCTCCCG ATGTAAAAGG	4380
	TACTGTTACT GTATATTCAT CTGACGTTAA ACCTGAAAAT CTACGCAATT TCTTTATTTC	4440
	TGTTTTACGT GCTAATAATT TTGATATGGT TGGTTCAATT CCTTCCATAA TTCAGAAGTA	4500
	TAATCCAAAC AATCAGGATT ATATTGATGA ATTGCCATCA TCTGATAATC AGGAATATGA	4560
	TGATAATTCC GCTCCTTCTG GTGGTTTCTT TGTTCCGCAA AATGATAATG TTACTCAAAC	4620
	TTTTAAAATT AATAACGTTC GGGCAAAGGA TTTAATACGA GTTGTCGAAT TGTTTGTAAA	4680
	GTCTAATACT TCTAAATCCT CAAATGTATT ATCTATTGAC GGCTCTAATC TATTAGTTGT	4740
	TAGTGCACCT AAAGATATTT TAGATAACCT TCCTCAATTC CTTTCTACTG TTGATTTGCC	4800
	AACTGACCAG ATATTGATTG AGGGTTTGAT ATTTGAGGTT CAGCAAGGTG ATGCTTTAGA	4860
	TTTTTCATTT GCTGCTGGCT CTCAGCGTGG CACTGTTGCA GGCGGTGTTA ATACTGACCG	4920
	CCTCACCTCT GTTTTATCTT CTGCTGGTGG TTCGTTCGGT ATTTTTAATG GCGATGTTTT	4980
	AGGGCTATCA CTTCGCGCAT TAAAGACTAA TAGCCATTCA AAAATATTGT CTGTGCCACG	5040
	TATTCTTACG CTTTCAGGTC AGAAGGGTTC TATCTCTGTT GGCCAGAATG TCCCTTTTAT	5100
	TACTGGTCGT GTGACTGGTG AATCTGCCAA TGTAAATAAT CCATTTCAGA CGATTGAGCG	5160
	TCAAAATGTA GGTATTTCCA TGAGCGTTTT TCCTGTTGCA ATGGCTGGCG GTAATATTGT	5220
	TCTGGATATT ACCAGCAAGG CCGATAGTTT GAGTTCTTCT ACTCAGGCAA GTGATGTTAT	5280
	TACTAATCAA AGAAGTATTG CTACAACGGT TAATTTGCGT GATGGACAGA CTCTTTTACT	5340
-	CGGTGGCCTC ACTGATTATA AAAACACTTC TCAAGATTCT GGCGTACCGT TCCTGTCTAA	5400
	AATCCCTTTA ATCGGCCTCC TGTTTAGCTC CCGCTCTGAT TCCAACGAGG AAAGCACGTT	5460
	ATACGTGCTC GTCAAAGCAA CCATAGTACG CGCCCTGTAG CGGCGCATTA AGCGCGGCGG	5520
	GTGTGGTGGT TACGCGCAGC GTGACCGCTA CACTTGCCAG CGCCCTAGCG CCCGCTCGTT	5580

THE STANDARD	5640
TCGCTTTCTT CCCTTCCTTT CTCGCCACGT TCGCCGGCTT TCCCCGTCAA GCTCTAAATC	5700
THE CONTROL TITA COUTT CONTINUE CONTINUES CONT	
TOTAL TOTAL AGTGGGCCAT CGCCCTGATA GACGGTTTTT CGCCCTTAN	5760
TOTAL CACGITATIT AATAGIGGAC TOTIGITAGA AACIGGAACA ACAGICAACO	5820
TOTAL CONTROL CTATTCTTTT GATTTATAAG GGATTTTGCC GATTTCGGAA CGACCATCAR	.5880
CCCCTCCTCC GCCAAACCAG CGTGGACCGC TTGCTGCAAC TOTOTOAGGG	5940
TO GOOGLE AAGGGGAATC AGCTGTTGCC CGTCTCGCTG GTGAAAAGAA AAACOACOCI	6000
ACCCARACT ACCCARACCG CCTCTCCCCG CGCGTTGGCC GATTCATTAA TGCAGGTGGC	6060
TOCCGACTEG AAAGCGGGCA GTGAGCGCAA CGCAATTAAT GTGAGTTAGO	6120
TOATTA GCCACCCCAG GCTTTACACT TTATGCTTCC GGCTCGTATG 11G1G1GGAA	6180
TOTALCCCC ATAACAATTT CACACGCGTC ACTTGGCACT GGCCGTCGTT TTACAACGTC	6240
TO CTCCCA ANACCCTGGC GTTACCCAAG CTTTGTACAT GGAGAAATA AAGIGAAACA	6300
ACCUSTATE CCACTGGGAC TOTTACCGTT ACTGTTTACC CCTGTGGCAA AAGUCTATGG	6360
TO STRUCTURE ACTICITICAGE GATCCGGAGC TGAAGGCGAT GACCCTGCTA AGGCTGCATT	6420
CACCCAAGTG CTACTGAGTA CATTGGCTAC GCTTGGGCTA TGGTAGTAGT	6480 6540
THE COTTOCT COTACCATAG GGATTAAATT ATTCAAAAAG TTTACGAGGA AGGCITOTA	6600
TARGE AACAGGCCG CACCGATCGC CCTTCCCAAC AGTTGCGCAG CUIGAAIGGC	6660
TOCCTOCT TECCTOCTT TECCGCACCA GAAGEGGTGC CGGAAAGUTG GETGGAGTG	6720
ACCCCGATAC GGTCGTCGTC CCCTCAAACT GGCAGATGCA CGGTACGAT	6780
ACACCAACGT AACCTATCCC ATTACGGTCA ATCCGCCGT1 IG11000A00	6840
TOTAL COCCUTTATA CTCGCTCACA TITAATGTTG ATGAAAGCTG GCIACAGGAA	6900
CAATTATITT TGATGGCGTT CCTATTGGTT AAAAAATGAG CIGATITAAC	6960
AAAAATTTAA CGCGAATTTT AACAAAATAT TAACGTTTAC AATTTAAATA TTTGCTTATA	7020
TATCAACCGG GGTACATRIG ATTOMOSTO	
TAGTTTTACG ATTACCGTTC ATCGATTCTC TTGTTTGCTC CAGACTCTCA GGCAATGACC	7140
TOTACATCTC TCAAAAATAG CTACCCTCTC CGGCAIIAAI IIAIGAGGAI	7200
GAACGGTTGA ATATCATATT GATGGTGATT TGACTGTCTC CGGCCTTTCT CACCCTTTTG	
GAACGGTTGA ATATCATATT CHARACTER CATTTAGAAAT ATATGAGGGT TCTAAAAATT AATCTTTACC TACACATTAC TCAGGCATTG CATTTAGACGCT CATAATGTTT	
AATCTTTACC TACACATTAC TOAGGGTTCTC CCGCAAAAGT ATTACAGGGT CATAATGTTT TTTATCCTTG CGTTGAAATA AAGGCTTCTC CCGCAAAAGT ATTACAGGGT CATAATGTTT	7.380
TTTATCCTTG CGTTGAAATA AAGGGTTOTO TTGGTACAAC CGATTTAGCT TTATGCTCTG AGGCTTTATT GCTTAATTTT GCTAATTCTT	7409
TGCCTTGCCT GTATGATTTA TTGGACGTT	

(2) INFORMATION FOR SEQ ID NO:5:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7294 bas pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: circular

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AATGCTACTA CTATTAGTAG	AATTGATGC	ACCTTTTCAC	CTCCCCCCCC	AAATGAAAAT	60
ATAGCTAAAC AGGTTATTGA	CCATTTGCGA	AATGTATCTA	ATGGTCAAAC	TAAATCTACT	120
CGTTCGCAGA ATTGGGAATC	AACTGTTACA	TGGAATGAAA	CTTCCAGACA	CCGTACTTTA	180
GTTGCATATT TAAAACATGT	TGAGCTACAG	CACCAGATTC	AGCAATTAAG	CTCTAAGCCA	240
TCTGCAAAAA TGACCTCTTA	TCAAAAGGAG	CAATTAAAGG	TACTCTCTAA	TCCTGACCTG	300
TTGGAGTTTG CTTCCGGTCT	GGTTCGCTTT	GAAGCTCGAA	TŢAAAACGCG	ATATTTGAAG	360
TCTTTCGGGC TTCCTCTTAA	TCTTTTTGAT	GCAATCCGCT	TTGCTTCTGA	CTATAATAGT	420
CAGGGTAAAG ACCTGATTTT	TGATTTATGG	TCATTCTCGT	TTTCTGAACT	GTTTAAAGCA	480
TTTGAGGGGG ATTCAATGAA	TATTTATGAC	GATTCCGCAG	TATTGGACGC	TATCCAGTCT	540
AAACATTTTA CTATTACCCC	CTCTGGCAAA	ACTTCTTTTG	CAAAAGCCTC	TCGCTATTTT	600
GGTTTTTATC GTCGTCTGGT	AAACGAGGGT	TATGATAGTG	TTGCTCTTAC	TATGCCTCGT	660
AATTCCTTTT GGCGTTATGT	ATCTGCATTA	GTTGAATGTG	GTATTCCTAA	ATCTCAACTG	720
ATGAATCTTT CTACCTGTAA	TAATGTTGTT	CCGTTAGTTC	GTTTTATTAA	CGTAGATTTT	780
TCTTCCCAAC GTCCTGACTG	GTATAATGAG	CCAGTTCTTA	AAATCGCATA	AGGTAATTCA	840
CAATGATTAA AGTTGAAATT	AAACCATCTC	AAGCCCAATT	TACTACTCGT	TCTGGTGTTT	900
CTCGTCAGGG CAAGCCTTAT	CACTGAATG .	AGCAGCTTTG	TTACGTTGAT '	TTGGGTAATG	960
AATATCCGGT TCTTGTCAAG	ATTACTCTTG	ATGAAGGTCA	GCCAGCCTAT	GCGCCTGCTC	1020
TGTACACCGT TCATCTGTCC	ICTTTCAAAG	TTGGTCAGTT	CCCTTCCCTT	ATGATTGACC	1080
GTCTGCGCCT CGTTCCGGCT	AAGTAACATG	GAGCAGGTCG	CGGATTTCGA	CACAATTTAT	1140
CAGGCGATGA TACAAATCTC	CGTTGTACTT	TGTTTCGCGC	TTGGTATAAT	CGCTGGGGGT	1200
CAAAGATGAG TGTTTTAGTG	TATTCTTTCG	CCTCTTTCGT	TTTAGGTTGG	TGCCTTCGTA	1260
GTGGCATTAC GTATTTTACC	CGTTTAATGG	AAACTTCCTC	ATGAAAAAGT	CTTTAGTCCT	1320
CAAAGCCTCT GTAGCCGTTG	CTACCCTCGT	TCCGATGCTG	TCTTTCGCTG	CTGAGGGTGA	1380
CGATCCCGCA AAAGCGGCCT	TTAACTCCCT	GCAAGCCTCA	GCGACCGAAT	ATATCGGTTA	1440
TGCGTGGGCG ATGGTTGTTG	TCATTGTCGG	CGCAACTATC	GGTATCAAGC	TGTTTAAGAA	1500
ATTCACCTCG AAAGCAAGCT		•			1560
TTTTTGGAGA TTTTCAACGT					1620
TATTCTCACT CCGCTGAAAC					1680

		- · mccTT	ACCCTAACTA '	TGAGGGTTGT	L/40
TTTACTAACG TCTGGAAAGA C	GACAAAACT	TTAGATUGII	AAACTCAGTG	TTACGGTACA	1800
	TOTA CTTTCT	ACTGGTGACG	AMMOIGHOU		1860
	TATCCCTGAA	AATGAGGGIG	GIGGOTOTO		1920
TO COMMOTICAL	CCCTCCCCGT	ACTAAACCIC	CIGNGINGOO		1980
TATE A COMMANDATE	CAACCCTCTC	GACGGCACII	AIGGGGGTG		2040
- acm ACCC	TTCTCTTGAG	GAGTCTCAGG	CICITATING		2100
	TACCCAGGGG	GCATTAACTG	TITATAGGGG		2160
A MACACITY A A	AACTTATTAC	CAGTACACTC	CIGINIONIO		2220
	TAAATTCAGA	GACTGCGCTT	ICCVIICIO		2280
mmmomca ATA	TCAAGGGGAA	TCGTCTGAGO	IGCOLOURIS		2340
comorrorror	TECTTETTGGT	GGCGGCTCTG	AGGGIGGIGG		2400
ACCCTCCCCC	CTCTGAGGGA	GGCGGTTCCG	GIGGIGGOID		2460
AMOAAAACAT	CCCAAACGCT	AATAAGGGGG	CIAIGAGGG		2520
THE THE CAST CTCTCA	CCCTAAAGGC	AAACTIGATI	CIGICOCINO		2580
AMAGETTECAT	TCCTGACGTT	TCCGGCCTTC	; CIAAIGGIAA	10000	2640
CHOCCTCTAA	TTCCCAAATC	GCTCAAGTC	, GIGNOGGIOL		2700
	ATATTTACC	T TCCCTCCCT	C WWICGGIION		2760
	ACCATATGA	A TITICTALL	G AllGIGHOLE	*	2820
	TCTTTTTATA	T GTTGCCACC	T TIMIGIALO.		2880
TA OTCCCTAA	TAAGGAGTC	T TAATCATGO	C WGITOTTT		2940
moomcccc	r TTCCTTCTG	G TAACTITGI	I CGGGTHIE		3000
TATTATIGCG TITCCICGGIA	3 ATAGCTATT	G CTATTICAL	T GITTOTTOO	A CCCTCTGACT	3060
GGCTTAACTC AATTCTTGT	G GGTTATCT	CT CTGATATIA	om mcccmcTTT	T TATGTTATTC	3120
GGCTTAACTC AATTCTTGTC TTGTTCAGGG TGTTCAGTT	A ATTCTCCC	GT CTAATGCG	CI ICCCIOILL	T TCTTATTTGG	3180
·	THE PROPERTY OF THE PARTY OF TH	TC ACGITAAA	ON WHITE		3240
	TO THE PROPERTY OF THE	TT CTAAGIGG	UM MALLELLE		3300
	TCACCATA	AA ATTGTAG	110 GG1001=-		3360
	A COTOCOGO	TAA GTUGGGA	3G1 100022		· ·
CTTGATTTAA GGCTTCAA CTTAGAATAC CGGATAAG	CC TTCTATA	TCT GATTIGO	ATC ACTCCGG	TAC TIGGTTTAAT	3480
CTTAGAATAC CGGATAAG TCCTACGATG AAAATAAA	AA CGGCTTG	CTT GTTCTCG	TTC ATTGGTT	TCT ACATGCTCGT	3540
TCCTACGATG AAAATAAA ACCCGTTCTT GGAATGAT	CAA GGAAAGA	CAG CCGATTA	TATE CTATTGT	TGA TAAACAGGCG	3600
ACCCGTTCTT GGAATGAT AAATTAGGAT GGGATAT	TAT CTTCCT	IGTI CAGGAC	COTO TOGACAO	AAT TACTTTACCT	3660
AAATTAGGAT GGGATAT CGTTCTGCAT TAGCTGA	ACA TGTTGT	TTAT TGTCGT	CGIO IGGADAC	GCC TAAATTACAT	3720
CGTTCTGCAT TAGCTGA TTTGTCGGTA CTTTATA	TTC TCTTAT	TACT GGCTCG	AAAA IGOOLO		

GTTGGCGTTG TTAAATATGG CGATTCTCAA TTAAGCCCTA CTGTTGAGCG TTGGCTTTAT	3780
ACTGGTAAGA ATTTGTATAA CGCATATGAT ACTAAACAGG CTTTTTCTAG TAATTATGAT	3840
TCCGGTGTTT ATTCTTATTT AACGCCTTAT TTATCACACG GTCGGTATTT CAAACCATTA	3900
AATTTAGGTC AGAAGATGAA GCTTACTAAA ATATATTTGA AAAAGTTTTC ACGCGTTCTT	-3960
TGTCTTGCGA TTGGATTTGC ATCAGCATTT ACATATAGTT ATATAACCCA ACCTAAGCCG	4020
GAGGTTAAAA AGGTAGTCTC TCAGACCTAT GATTTTGATA AATTCACTAT TGACTCTTCT	4080
CAGCGTCTTA ATCTAAGCTA TCGCTATGTT TTCAAGGATT CTAAGGGAAA ATTAATTAAT	4140
AGCGACGATT TACAGAAGCA AGGTTATTCA CTCACATATA TTGATTTATG TACTGTTTCC	4200
ATTAAAAAGG TAATTCAAAT GAAATTGTTA AATGTAATTA ATTTTGTTTT CTTGATGTTT	4260
GTTTCATCAT CTTCTTTTGC TCAGGTAATT GAAATGAATA ATTCGCCTCT GCGCGATTTT	4320
GTAACTTGGT ATTCAAAGCA ATCAGGCGAA TCCGTTATTG TTTCTCCCGA TGTAAAAGGT	4380
ACTGTTACTG TATATTCATC TGACGTTAAA CCTGAAAATC TACGCAATTT CTTTATTTCT	4440
GTTTTACGTG CTAATAATTT TGATATGGTT GGTTCAATTC CTTCCATTAT TTAGAAGTAT	4500
AATCCAAACA ATCAGGATTA TATTGATGAA TTGCCATCAT CTGATAATCA GGAATATGAT	4560
GATAATTCCG CTCCTTCTGG TGGTTTCTTT GTTCCGCAAA ATGATAATGT TACTCAAACT	4620
TTTAAAATTA ATAACGTTCG GGCAAAGGAT TTAATACGAG TTGTCGAATT GTTTGTAAAG	4680
TCTAATACTT CTAAATCCTC AAATGTATTA TCTATTGACG GCTCTAATCT ATTAGTTGTT	4740
AGTGCACCTA AAGATATTTT AGATAACCTT CCTCAATTCC TTTCTACTGT TGATTTGCCA	4800
ACTGACCAGA TATTGATTGA GGGTTTGATA TTTGAGGTTC AGCAAGGTGA TGCTTTAGAT	4860
TTTTCATTTG CTGCTGGCTC TCAGCGTGGC ACTGTTGCAG GCGGTGTTAA TACTGACCGC	4920
CTCACCTCTG TTTTATCTTC TGCTGGTGGT TCGTTCGGTA TTTTTAATGG CGATGTTTTA	4980
GGGCTATCAG TTCGCGCATT AAAGACTAAT AGCCATTCAA AAATATTGTC TGTGCCACGT	5040
ATTCTTACGC TTTCAGGTCA GAAGGGTTCT ATCTCTGTTG GCCAGAATGT CCCTTTTATT	5100
ACTGGTCGTG TGACTGGTGA ATCTGCCAAT GTAAATAATC CATTTCAGAC GATTGAGCGT	5160
CAAAATGTAG GTATTTCCAT GAGCGTTTTT CCTGTTGCAA TGGCTGGCGG TAATATTGTT	5220
CTGGATATTA CCAGCAAGGC CGATAGTTTG AGTTCTTCTA CTCAGGCAAG TGATGTTATT	5280
ACTAATCAAA GAAGTATTGC TACAACGGTT AATTTGCGTG ATGGACAGAC TCTTTTACTC	5340
GGTGGCCTCA CTGATTATAA AAACACTTCT CAAGATTCTG GCGTACCGTT CCTGTCTAAA	5400
ATCCCTTTAA TCGGCCTCCT GTTTAGCTCC CGCTCTGATT CCAACGAGGA AAGCACGTTA	5460
TACGTGCTCG TCAAAGCAAC CATAGTACGC GCCCTGTAGC GGCGCATTAA GCGCGGCGGG	5520
TGTGGTGGTT ACGCGCAGCG TGACCGCTAC ACTTGCCAGC GCCCTAGCGC CCGCTCCTTT	5580
CGCTTTCTTC CCTTCCTTTC TCGCCACGTT CGCCGGCTTT CCCCGTCAAG CTCTAAATCG	5640
GGGGCTCCCT TTAGGGTTCC GATTTAGTGC TTTACGGCAC CTCGACCCCA AAAAACTTGA	57.00
TTTGGGTGAT GGTTCACGTA GTGGGCCATC GCCCTGATAG ACGGTTTTTC GCCCTTTGAC	5760

STIGGAGICC ACGIICITTA ATAGIGGACI CITGIICCAA ACIGGAACAA CACICAACCC	7020
TATCTCGGGC TATTCTTTTG ATTTATAAGG GATTTTGCCG ATTTCGGAAC CACCATCAAA	5880
TATCTCGGGC TATTCTTTTG ATTIATAAGG GATTTCGGA TECTCGAACT CTCTCAGGGC	5940
CAGGATTTTC GCCTGCTGGG GCAAACCAGC GTGGACCGCT TGCTGCAACT CTCTCAGGGC	6000
CAGGCGGTGA AGGGCAATCA GCTGTTGCCC GTCTCGCTGG TGAAAAGAAA AACCACGCTG	6060
GCGCCCAATA CGCAAACCGC CTCTCCCCGC GCGTTGGCCG ATTCATTAAT GCACGTOTAT	
CGACAGGTTT CCCGACTGGA AAGCGGGCAG TGAGCGCAAC GCAATTAATG TGAGTTAGCT	6120
CACTCATTAG GCACCCCAGG CTTTACACTT TATGCTTCCG GCTCGTATGT TGTGTGGAAT	6180
TGTGAGCGGA TAACAATTTC ACACAGGAAA CAGCTATGAC CAGGATGTAC GAATTCGCAG	6240
TGTGAGGGGA TAACAATTTO NOMONOOTOO	6300
GTAGGAGAGC TCGGCGGATC CGAGGCTGAA COCCATACATT GGGCTATGGT AGTAGTTATA	6360
AGTTTACAGG CAAGTGCTAC TGAGTACATT GGGTACGGT CGAGGAAGGC TTCTTAACCA	6420
GTTGGTGCTA CCATAGGGAT TAAATTATTC AAAAAGTTTA CGAGCAAGGC TTCTTAACCA	6480
GCTGGCGTAA TAGCGAAGAG GCCCGCACCG ATCGCCCTTC CCAACAGTTG CGCAGCCTGA	6540
ATCCCCAATG GCGCTTTGCC TGGTTTCCGG CACCAGAAGC GGTGCCGGAA AGCTGGGTGC	6600
ACTGCGATCT TCCTGAGGCC GATACGGTCG TCGTCCCCTC AAACIGGGAA ATGGACA	
ACGATGGGCC CATCTACACC AACGTAACCT ATCCCATTAC GGTCAATCCG CCGTTTGTTC	6660
CCACGGAGAA TCCGACGGT TGTTACTCGC TCACATTTAA TGTTGATGAA AGCTGGCTAC	6720
AGGAAGGCCA GACGCGAATT ATTTTTGATG GCGTTCCTAT TGGTTAAAAA ATGAGCTGAT	6780
TTAACAAAAA TTTAACGCGA ATTTTAACAA AATATTAACG TTTACAATTT AAATATTTGC	6840
TTATACAATC TTCCTGTTTT TGGGGCTTTT CTGATTATCA ACCGGGGTAC ATATGATTGA	6900
CATGCTAGTT TTACGATTAC CGTTCATCGA TTCTCTTGTT TGCTCCAGAC TCTCAGGCAA	6960
TGACCTGATA GCCTTTGTAG ATCTCTCAAA AATAGCTACC CTCTCCGGCA TTAATTTATC	7020
TGACCTGATA GCCTTTGTAG ATCTCTCAAA AATAGGTACT CTCTCCGGGC TTTCTCAGCC	7080
AGCTAGAACG GTTGAATATC ATATTGATGG TGATTTGACT GTCTCCGGCC TTTCTCACCC	7140
TITIGAATCT TTACCTACAC ATTACTCAGG CATTGCATTT AAAATATATG AGGGTTCTAA	7200
AAATTTTTAT CCTTGCGTTG AAATAAAGGC TTCTCCCGCA AAAGTATTAC AGGGTCATAA	7260
TGTTTTTGGT ACAACCGATT TAGCTTTATG CTCTGAGGCT TTATTGCTTA ATTTTGCTAA	7294
TTCTTTGCCT TGCCTGTATG ATTTATTGGA CGTT	1234

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:

 (A) LENGTH: 7394 base pairs

 (B) TYPE: nucleic acid

 (C) STRANDEDNESS: both

 (D) TOPOLOGY: circular
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

ATAGCTAAAC AGGTTATTGA CCATTTGCGA AATGTATCTA ATGGTCAAAC TAAATCTACT	120
CGTTCGCAGA ATTGGGAATC AACTGTTACA TGGAATGAAA CTTCCAGACA CCGTACTTTA	180
GTTGCATATT TAAAACATGT TGAGCTACAG CACCAGATTC AGCAATTAAG CTCTAAGCCA	240
TCTGCAAAAA TGACCTCTTA TCAAAAGGAG CAATTAAAGG TACTCTCTAA TCCTGACCTG	300
TTGGAGTTTG CTTCCGGTCT GGTTCGCTTT GAAGCTCGAA TTAAAACGCG ATATTTGAAG	360
TCTTTCGGGC TTCCTCTTAA TCTTTTTGAT GCAATCCGCT TTGCTTCTGA CTATAATAGT	420
CAGGGTAAAG ACCTGATTTT TGATTTATGG TCATTCTCGT TTTCTGAACT GTTTAAAGCA	480
TTTGAGGGGG ATTCAATGAA TATTTATGAC GATTCCGCAG TATTGGACGC TATCCAGTCT	540
AAACATTTTA CTATTACCCC CTCTGGCAAA ACTTCTTTTG CAAAAGCCTC TCGCTATTTT	600
GGTTTTTATC GTCGTCTGGT AAACGAGGGT TATGATAGTG TTGCTCTTAC TATGCCTCGT	660
AATTCCTTTT GGCGTTATGT ATCTGCATTA GTTGAATGTG GTATTCCTAA ATCTCAACTG	720
ATGAATCTTT CTACCTGTAA TAATGTTGTT CCGTTAGTTC GTTTTATTAA CGTAGATTTT	780
TCTTCCCAAC GTCCTGACTG GTATAATGAG CCAGTTCTTA AAATCGCATA AGGTAATTCA	840
CAATGATTAA AGTTGAAATT AAACCATCTC AAGCCCAATT TACTACTCGT TCTGGTGTTT	900
CTCGTCAGGG CAAGCCTTAT TCACTGAATG AGCAGCTTTG TTACGTTGAT TTGGGTAATG	960
AATATCCGGT TCTTGTCAAG ATTACTCTTG ATGAAGGTCA GCCAGCCTAT GCGCCTGGTC	1020
TGTACACCGT TCATCTGTCC TCTTTCAAAG TTGGTCAGTT CGGTTCCCTT ATGATTGACC	1080
GTCTGCGCCT CGTTCCGGCT AAGTAACATG GAGCAGGTCG CGGATTTCGA CACAATTTAT	1140
CAGGCGATGA TACAAATCTC CGTTGTACTT TGTTTCGCGC TTGGTATAAT CGCTGGGGGT	1200
CAAAGATGAG TGTTTTAGTG TATTCTTTCG CCTCTTTCGT TTTAGGTTGG TGCCTTCGTA	1260
GTGGCATTAC GTATTTTACC CGTTTAATGG AAACTTCCTC ATGAAAAAGT CTTTAGTCCT	1320
CAAAGCCTCT GTAGCCGTTG CTACCCTCGT TCCGATGCTG TCTTTCGCTG CTGAGGGTGA	1380
CGATCCCGCA AAAGCGGCCT TTAACTCCCT GCAAGCCTCA GCGACCGAAT ATATCGGTTA	1440
TGCGTGGGCG ATGGTTGTTG TCATTGTCGG CGCAACTATC GGTATCAAGC TGTTTAAGAA	1500
ATTCACCTCG AAAGCAAGCT GATAAACCGA TACAATTAAA GGCTCCTTTT GGAGCCTTTT	1560
TTTTGGAGA TTTTGAACGT GAAAAATTA TTATTCGCAA TTCCTTTAGT TGTTCCTTTC	1620
FATTCTCACT CCGCTGAAAC TGTTGAAAGT TGTTTAGCAA AACCCCATAC AGAAAATTCA	1680
TTTACTAACG TCTGGAAAGA CGACAAAACT TTAGATCGTT ACGCTAACTA TGAGGGTTGT	1740
CTGTGGAATG CTACAGGCGT TGTAGTTTGT ACTGGTGACG AAACTCAGTG TTACGGTACA	1800
TGGGTTCCTA TTGGGCTTGC TATCCCTGAA AATGAGGGTG GTGGCTCTGA GGGTGGCGGT	1860
TCTGAGGGTG GCGGTTCTGA GGGTGGCGGT ACTAAACCTC CTGAGTACGG TGATACACCT	1920
ATTCCGGGCT ATACTTATAT CAACCCTCTC GACGGCACTT ATCCGCCTGG TACTGAGCAA	1980
AACCCCGCTA ATCCTAATCC TTCTCTTGAG GAGTCTCAGC CTCTTAATAC TTTCATGTTT	2040
GAGATAATA CCTTCCGAAA TAGGCAGGGG GCATTAACTG TTTATACGGG GACTGTTACT	2100

THE CASE CECTATCATC AAAAGCCATG	2160
CAAGGCACTG ACCCCGTTAA AACTTATTAC CAGTACACTC CTGTATCATC AAAAGCCATG	2220
CAAGGCACTG ACCCCGTTAA AACTTATTATGAA TATGACGCTT ACTGGAACGG TAAATTCAGA GACTGCGCTT TCCATTCTGG CTTTAATGAA	2280
TATGACGCTT ACTGGAACG TAAATTOTAG GATCCATTCG TTTGTGAATA TCAAGGCCAA TCGTCTGACC TGCCTCAACC TCCTGTCAAT GATCCATTCG ACCCTGGTGG CTCTGAGGGT	2340
GATCCATTCG TITGIGARIA TOMICO GEOGGETETG AGGGTGGTGG CTCTGAGGGT GCTGGCGGCG GCTCTGGTGG TGGTTCTGGT GGCGGCTCTG AGGGTGGTG CTCTGAGGGT	2400
GCTGGCGGCG GCTCTGGTGG TGGTTCCGGT GGCGGTTCTG AGGGTGGCGG CTCTGAGGGA GGCGGTTCCG GTGGTGGCTC TGGTTCCGGT	-2460
GGCGGTTCTG AGGGTGGCGG GTGTAGACGCT AATAAGGGGG CTATGACCGA AAATGCCGAT GATTTTGATT ATGAAAAGAT GGCAAACGCT AATAAGGGGG CTATGACCGA AAATGCCGAT	2520
GAAAACGCGC TACAGTCTGA CGCTAAAGGC AAACTTGATT CTGTCGCTAC TGATTACGGT	2580
GAAAACGCGC TACAGTCTGA CGGTMATCACTT TCCGGCCTTG CTAATGGTAA TGGTGCTACT GCTGCTATCG ATGGTTCAT TGGTGACGTT TCCGGCCTTG CTAATGGTAA TGGTGCTACT	2640
GCTGCTATCG ATGGILLGAT TGGTGATGGCTGA GCTGAGGTGG GTGACGGTGA TAATTCACCT GGTGATTTTG CTGGCTCTAA TTCCCAAATG GCTCAAGTCG GTGACGGTTGA ATGTCGCCCT	27.00
GGTGATTITG CTGGCTCIAA TTOOSIZED TCGCTCCCTC AATCGGTTGA ATGTCGCCCT TTAATGAATA ATTTCCGTCA ATATTTACCT TCCCTCCCTC AATCGGTTGA ATGTCGCCCT TTAATGAATA ATTTCCGTCA ATATTTCCTTATTCC ATTCTGACAA AATAAACTTA	2760
TTAATGAATA ATTTCGGTGA ATATTCTATTG ATTGTGACAA AATAAACTTA TTTGTCTTTA GCGCTGGTAA ACCATATGAA TTTTCTATTG ATTGTGACAA AATAAACTTA	2820
TTTGTCTTTA GCGCTGGTAA NOOTTTATATAT GTTGCCACCT TTATGTATGT ATTTTCTACG TTCCGTGGTG TCTTTGCGTT TCTTTTATAT GTTGCCACCT TTATGTATGT ATTTTCTACGT	2880
TTCCGTGGTG TCTTTGCGTA TAAGGAGTCT TAATCATGCC AGTTCTTTTG GGTATTCCGT TTTGCTAACA TACTGCGTAA TAAGGAGTCT TAATCATGCC AGTTCTTTTC GGCTATCTC	2940
TATTATTGCG TTTCCTCGGT TTCCTTCTGG TAACTTTGTT CGGCTATCTG CTTACTTTTC TATTATTGCG TTTCCTCGGT TTCCTTCTGG TAACTTTGTT CGGCTATCTG CTTACTTTTC	3000
TATTATTGCG TITCCICGGI TITOTTCATT CTATTTCATT GTTTCTTGCT CTTATTATTG TTAAAAAGGG CTTCGGTAAG ATAGCTATTG CTATTTAG CGCTCAATTA CCCTCTGACT	3060
GGCTTAACTC AATTCTTGTG GGTTATCTCT CTGATATTAG CGCTCAATTA CCCTCTGACT	3120
TCTCTGTAAA GGCTGCTATT TTCATTTTTG ACGTTAAACA AAAAATCGTT TCTTATTTGG TCTCTGTAAA GGCTGCTATT TTCATTTTTG ACGTTAAACA AAAAATCGTT TCTTATTTGG	3180
TCTCTGTAAA GGCTGCTATT TTCATTTTTG ROOTTATTTTG GTAACTGGCA AATTAGGCTC TGGAAAGACG ATTGGGATAA ATAATATGGC TGTTTATTTT GTAACTGGCA AATTAGGCTC TGGAAAGACG	3240
ATTGGGATAA ATAATATGGC TGTTTATTTT GTATGTAGCTG GGTGCAAAAT AGCAACTAAT CTCGTTAGCG TTGGTAAGAT TTAGGATAAA ATTGTAGCTG GGTGCAAAAC GCGTCGCGTT	3300
CTCGTTAGCG TTGGTAAGAT TTAGGATAAA ATTOTTO CTATTGGGGG CGGTAATGAT CTTGATTTAA GGCTTCAAAA CCTCCCGCAA GTCGGGAGGT TCGCTAAAAC GCCTCGCGTT	3360
CTTGATTTAA GGCTTCAAAA CCTCCCGCAA GTCGCCGTTG CTATTGGGCG CGGTAATGAT CTTAGAATAC CGGATAAGCC TTCTATATCT GATTTGCTTG ACTCCGGTAC TTGGTTTAAT	3420
COCCUTCUTT GTTCTCGATG AGIGGOUST	3480
COLLACACAC CCGATIALIG ALLOCALA	3540
TOTAL CALGACTER CALCACTER	•
TOTTOTTAT TGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG	
CGTTCTGCAT TAGCTGAACA IGITGTTATT TTTGTCGGTA CTTTATATTC TCTTATTACT GGCTCGAAAA TGCCTCTGCC TAAATTACAT	3720
TAAGGUULA GIGI	
COCATATCAT ACTAAACAGG CIIII	
AACCCCTTAT TIAICACACO GIO	
COMPACTADA AIAILIGA 14	
ATCACCATTT ACATATACTI ATTACA	
TCACACCTAT GATILIGATA INILIZATA	
GAGGTTAAAA AGGTAGTCIC TOAGAGGTT TICAAGGATT CTAAGGGAAA ATTAATTAA	T 414(
ONCCOTOTIA ALGINOUS TOTAL	

					TACTGTTTCC	4200
					TCTTGATGTT	4260
					TGCGCGATTT	4320
TGTAACTTGG	TATTCAAAGC	AATCAGGCGA	ATCCGTTAT	GTITCTCCC	G ATGTAAAAGG	4380
TACTGTTACT	GTATATTCAT	CTGACGTTAA	ACCTGAAAAT	CTACGCAAT	TCTTTATTTC	4440
TGTTTTACGT	GCTAATAATT	TTGATATGGT	TGGTTCAATT	CCTTCCATAA	TTCAGAAGTA	4500
TAATCCAAAC	AATCAGGATT	ATATTGATGA	ATTGCCATCA	TCTGATAATC	AGGAATATGA	4560
TGATAATTCC (CTCCTTCTG	GTGGTTTCTT	TGTTCCGCAA	AATGATAATG	TTACTCAAAC	4620
TTTTAAAATT A	AATAACGTTC	GGGCAAAGGA	TTTAATACGA	GTTGTCGAAT	TGTTTGTAAA	4680
GTCTAATACT T						4740
TAGTGCACCT A						4800
AACTGACCAG A			A CONTRACTOR OF THE CONTRACTOR			4860
TTTTTCATTT G						4920
CCTCACCTCT G	TTTTATCTT	CTGCTGGTGG	TTCGTTCGGT	ATTTTTAATG	GCGATGTTTT	4980
AGGGCTATCA G	TTCGCGCAT	TAAAGACTAA	TAGCCATTCA	AAAATATTGT	CTGTGCCACG	5040
TATTCTTACG C						5100
TACTGGTCGT G						5160
TCAAAATGTA G	GTATTTCCA	TGAGCGTTTT	TCCTGTTGCA	ATGGCTGGCG	GTAATATTGT	5220
TCTGGATATT A	CCAGCAAGG	CCGATAGTTT	GAGTTCTTCT	ACTCAGGCAA	GTGATGTTAT	5280
TACTAATCAA A	GAAGTATTG	CTACAACGGT	TAATTTGCGT	GATGGACAGA	CTCTTTTACT	5340
CGGTGGCCTC A	CTGATTATA	AAAACACTTC	TCAAGATTCT	GGCGTACCGT	TCCTGTCTAA	5400
AATCCCTTTA A	TCGGCCTCC	TGTTTAGCTC	CCGCTCTGAT	TCCAACGAGG	AAAGCACGTT	5460
ATACGTGCTC G	TCAAAGCAA	CCATAGTACG	CGCCCTGTAG	CGGCGCATTA	AGCGCGGCGG	5520
GTGTGGTGGT T	ACGCGCAGC	GTGACCGCTA	CACTTGCCAG	CGCCCTAGCG	CCCGCTCCTT	5580
TCGCTTTCTT C	CCTTCCTTT	CTCGCCACGT	TCGCCGGCTT	TCCCCGTCAA	GCTCTAAATC	5640
GGGGGCTCCC T	TTAGGGTTC	CGATTTAGTG	CTTTACGGCA	CCTCGACCCC	AAAAAACTTG	57 0 0
ATTTGGGTGA T	GGTTCACGT	AGTGGGCCAT	CGCCCTGATA	GACGGTTTTT	CGCCCTTTGA	5760
CGTTGGAGTC C	ACCTTCTTT	AATAGTGGAC	TCTTGTTCCA	AACTGGAACA	ACACTCAACC	5820
CTATCTCGGG C	TATTCTTTT	GATTTATAAG	GGATTTTGCC	GATTTCGGAA	CCACCATCAA	. 5880
ACAGGATTTT C	CCCTCCTCC	GGCAAACCAG	CGTGGACCGC	TTGCTGCAAC	TCTCTCAGGG	5940
					AAACCACCCT	6000
GGCGCCCAAT A			•	• ,		6060
					GTGAGTTAGC	6120
			_		TTGTGTGGAA	6180

CGCCGTCGTT TTACAACGTC	6240
TTGTGAGCGG ATAACAATTT CACACGCGTC ACTTGGCACT GGCCGTCGTT TTACAACGTC	6300
	6360
	6420
	6480
AAAMTO 12(1) A LUU 14	6540
AAGTGCTACT GAGTACATTG GCTACGCTTG GGGTACGCTTG GGGTATGCGAAGAG TAGCGAAGAG CATAGGGATT AAATTATTCA AAAAGTTTAC GAGCAAGGCT TCTTAAGCAA TAGCGAAGAG CATAGGGATT AAATTATTCA AAAAGTTTAC GAGCAAGGCT ATGGCGAATG GCGCTTTGCC	6600
CATAGGGATT AAATTATTCA AAAAGTTTAC GACCAGCCTGA ATGGCGAATG GCGCTTTGCC GCCCGCACCG ATCGCCCTTC CCAACAGTTG CGCAGCCTGA ATGGCGAATG GCGCTTTGCC TGGTTTCCGG CACCAGAAGC GGTGCCGGAA AGCTGGCTGG AGTGCGATCT TCCTGAGGCC TGGTTTCCGG CACCAGAAGC GGTGCCGGAA AGCTGGCTT ACGATGCGCC CATCTACACC	6660
	6720
GATACGGTCG TCGTCCCCTC AAACTGGCAG ATGONOOGAGAA TCCGACGGGT AACGTAACCT ATCCCATTAC GGTCAATCCG CCGTTTGTTC CCACGGAGAA TCCGACGGGT AACGTAACCT ATCCCATTAC AGGAAGGCCA GACGCGAATT	6780
- COMPARION ALICITIES TO THE PROPERTY OF THE P	6840
TO STORY A A A A A A I GAGGIGAL ALL	6900
TALLES TALLES	6960
ACCOCCTAC ATAIGALIGA	7020
- amaca ca	7080
	7140
	7200
	7260 7320
A OTT A TOTAL OF A GALLANDE	7380
AAATAAAGGC TTCTCCCGCA AAAGTATTAC ACCOUNTAGE TAGCTTTATG CTCTGAGGCT TTATTGCTTA ATTTTGCTAA TTCTTTGCCT TGCCTGTATG	7394
	1377
ATTTATTGGA CGTT	

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7: GATCCTAGGC TGAAGGCGAT GACCCTGCTA AGGCTGC

(2) INFORMATION FOR SEQ ID NO:8:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 35 base pairs (B) TYPE: nucleic acid

(C) STRANDEDNESS: single (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:					
ATTCAATAGT TTACAGGCAA GTGCTACTGA GTACA	٠				3.5
(2) INFORMATION FOR SEQ ID NO:9:		. • •			
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 35 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	*				(1)
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:			•		
TTGGCTACGC TTGGGCTATG GTAGTAGTTA TAGTT	•••	· p	÷	* · ·	35
(2) INFORMATION FOR SEQ ID NO:10:					
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 35 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear			i		
	•				
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:	*		-	*	
GGTGCTACCA TAGGGATTAA ATTATTCAAA AAGTT					35
, , , , , , , , , , , , , , , , , , ,					
(2) INFORMATION FOR SEQ ID NO:11:					
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 18 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 				-	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:			• *		1.0
TACGAGCAAG GCTTCTTA		•			18
(2) INFORMATION FOR SEQ ID NO:12:					
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear					·
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:					
AGCTTAAGAA GCCTTGCTCG TAAACTTTTT GAATAATTT					39

(2) INFORMATION FOR SEQ ID NO:13:			
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 36 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear			
			7 · · · · · · · · · · · · · · · · · · ·
(xi) SEQUENCE DESCRIPTION: SEQ ID	NO:13:		
AATCCCTATG GTAGCACCAA CTATAACTAC TACCA	T	*	36
AATCCCTAIG GIAGOACOLLI		. * .	*
(2) INFORMATION FOR SEQ ID NO:14:	* 19		
(1) SEQUENCE CHARACTERISTICS: (A) LENGTH: 35 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear			
	* *		
(x1) SEQUENCE DESCRIPTION: SEQ ID	NO:14:		35
AGCCCAAGCG TAGCCAATGT ACTCAGTAGC ACTTG		* *	
(2) INFORMATION FOR SEQ ID NO:15:			
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 34 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear			*
(xi) SEQUENCE DESCRIPTION: SEQ ID	NO:15:		34
CCTGTAAACT ATTGAATGCA GCCTTAGCAG GGTC		1	
(2) INFORMATION FOR SEQ ID NO:16:	*		
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 16 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear			
*			
(xi) SEQUENCE DESCRIPTION: SEQ I	D NO:16:		1
ATCGCCTTCA GCCTAG			. *
(2) INFORMATION FOR SEQ ID NO:17:	(f)		. *
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 27 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: lin ar	*(

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:			;
CTCGAATTCG TACATCCTGG TCATAGC	T.	• . •	2
(2) INFORMATION FOR SEQ ID NO:18:		•	٠
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 21 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	*		
(x1) SEQUENCE DESCRIPTION: SEQ ID NO:18:			
CATTITIGCA GATGGCTTAG A		•	21
(2) INFORMATION FOR SEQ ID NO:19:			
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 18 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear			
(x1) SEQUENCE DESCRIPTION: SEQ ID NO:19:		•	
TAGCATTAAC GTCCAATA			18
(2) INFORMATION FOR SEQ ID NO:20:			
(1) SEQUENCE CHARACTERISTICS: (A) LENGTH: 26 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear			
		. *	
(x1) SEQUENCE DESCRIPTION: SEQ ID NO:20:	•		
ATATATTTA GTAAGCTTCA TCTTCT			26
(2) INFORMATION FOR SEQ ID NO:21:	•		
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 23 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear			
			*

23,

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

	:		
(2·)	INFORMATION FOR SEQ ID NO:22:		
X	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 35 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear		, ,
	(D) TOPOLOGY: Princes		
		- 20.00	
0	(xi) SEQUENCE DESCRIPTION: SEQ II		35
GCG	GGCCTCT TCGCTATTGC TTAAGAAGCC TTG	CT	8
(2)	INFORMATION FOR SEQ ID NO:23:		
nd pa	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 48 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear		*
	(xi) SEQUENCE DESCRIPTION: SEQ II	D NO:23:	48
TTC	AGCCTAG GATCCGCCGA GCTCTCCTAC CTG	CGAATTC GTACATCC	
٠.	70.24		
(2)	INFORMATION FOR SEQ ID NO:24:		
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANNEDNESS: single (D) TOPOLOGY: linear		1
	* *		
	(xi) SEQUENCE DESCRIPTION: SEQ I	ID NO:24:	24
TG	GATTATAC TICTAAATAA TGGA		•
(2) INFORMATION FOR SEQ ID NO:25:		
×.	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 36 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear		*
•			٠,
	(xi) SEQUENCE DESCRIPTION: SEQ		36
T	AACACTCAT TCCGGATGGA ATTCTGGAGT C	TGGGT	
(2) INFORMATION FOR SEQ ID NO:26:		*
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 22 base pair (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear		

TCTAGAACGC GTC

(x1) SEQUENCE DESCRIPTION: SEQ ID	NO:26:	
AATTCGCCAA GGAGACAGTC AT		22
(2) INFORMATION FOR SEQ ID NO:27:	γ.	-
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 39 base pairs	*	
(B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear		* 400
(xi) SEQUENCE DESCRIPTION: SEQ ID	NO:27:	· .•
AATGAAATAC CTATTGCCTA CGGCAGCCGC TGGATT	IGTT	39
(2) INFORMATION FOR SEQ ID NO:28:		
(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 39 base pairs(B) TYPE: nucleic acid		4.5 4.
(C) STRANDEDNESS: single (D) TOPOLOGY: linear		•
		<i>:</i>
(xi) SEQUENCE DESCRIPTION: SEQ ID NO	0:28:	
ATTACTCGCT GCCCAACCAG CCATGGCCGA GCTCGTC	GAT	. 39
(2) INFORMATION FOR SEQ ID NO:29:		
(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 39 base pairs(B) TYPE: nucleic acid		
(B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear		
,	*	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO):29:	
GACCCAGACT CCAGATATCC AACAGGAATG AGTGTTA	AAT	39
(2) INFORMATION FOR SEQ ID NO:30:		
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 13 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single		
(D) TOPOLOGY: linear		
(xi) SEQUENCE DESCRIPTION: SEQ ID NO	0:30:	

		N FOR SEQ						=	
(i	(A) (B)	ENCE CHARAC' LENGTH: 35 TYPE: nucl STRANDEDNE TOPOLOGY:	eic acid SS: single		*	:			
ومحل تسودان				o ID NO	· 11 ·	چانسان سال ما		i tali	
		ENCE DESCRI				30		,	3.5
ACGTGAC	GCG TT	CTAGAATT AA	CACTCATT (CCTGT					
(2) INF	ORMATI	ON FOR SEQ	ID NO:32:			- *		•	0
(i	(A) (B)	ENCE CHARAC LENGTH: 39 TYPE: nucl STRANDEDNE TOPOLOGY:	eic acid SS: single						
			•		, Ġ		- X		
		ENCE DESCRI				Y			39
TGGATAT	CTG GA	GTCTGGGT CA	TCACGAGC	TCGGCCA'	TG'	•			. . .
*			* 4	÷		÷	5	•	*
(2) IN	FORMATI	ON FOR SEQ	ID NO:33:			· .			
(:	(A) (B)	JENCE CHARAC LENGTH: 39 TYPE: nucl STRANDEDNI TOPOLOGY:	leic acid ESS: singl			•	0.		
			- 1	3		*			
(x	i) SEQ	UENCE DESCR	IPTION: SE	Q ID NO):33:	*			39
GCTGGT	TGGG C	AGCGAGTAA T	AACAATCCA	GCGGCTC	GCC 1.	·		T.	
		ION FOR SEQ							*
•	(A (I	QUENCE CHARA LENGTH: 3 TYPE: nuc STRANDEDM TOPOLOGY	cleic acid		<i>*</i>		*		
						,		** ,	
. (xi) SE	QUENCE DESC	RIPTION: S	SEQ ID	vo:34:	• • •			3
GTAGG	CAATA	GGTATTTCAT	TATGACTGT	CTTGG	CG	# # # # # # # # # # # # # # # # # # #			
(2) 1	INFORMA	TION FOR SE	Q ID NO:3	5:	*				
		QUENCE CHAR (A) LENGTH: (B) TYPE: DI (C) STRANDE (D) TOPOLOG	ucleic aci DNESS: sir	d		,	. •		4

(xi) SEQUENCE DESCRIPTION: SEQ ID NO	0:35:	* *	. "	. *.		
TGACTGTCTC CTTGGCGTGT GAAATTGTTA	* . *				•	3
(2) INFORMATION FOR SEQ ID NO:36:						
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 36 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear					:	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO	:36:		<i>;</i> •			
						36
TAACACTCAT TCCGGATGGA ATTCTGGAGT CTGGGT		-				,
(2) INFORMATION FOR SEQ ID NO:37:			•			
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 25 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	.* 9 *					
		3				,
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:	37:			*		
CAATTTTATC CTAAATCTTA CCAAC	•				•	25
(2) INFORMATION FOR SEQ ID NO:38:	•					
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 21 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear		•				
	• 0					
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:	38:					
CATITITGCA GATGGCTTAG A	• • • •				•	21
(2) INFORMATION FOR SEQ ID NO:39:						
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 21 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear		e* *		. · .		
and the second of the second o	•			·		
(x1) SEQUENCE DESCRIPTION: SEQ ID NO	: 39 :					

. 36

(2) INFORMATION FOR SEQ ID NO:40:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 18 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(xi) SEQUENCE DESCRIPTION: SEQ ID) NO:40:
TAGCATTAAC GTCCAATA	18
TAGGATTAAG GTOOMMAN	
(2) INFORMATION FOR SEQ ID NO:41:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 43 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(xi) SEQUENCE DESCRIPTION: SEQ ID) NO:41:
AAACGACGGC CAGTGCCAAG TGACGCGTGT GAAA	
AAACGACGGC CAGIGCCAAG IGAGGGGIGI CILLA	
(2) INFORMATION FOR SEQ ID NO:42:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 43 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(xi) SEQUENCE DESCRIPTION: SEQ I	D NO:42:
GGCGAAAGGG AATTCTGCAA GGCGATTAAG CTT	GGGTAAC GCC 43
(2) INFORMATION FOR SEQ ID NO:43:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 36 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single	

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:43: GGCGTTACCC AAGCTTTGTA CATGGAGAAA ATAAAG

38

(2) INFORMATION FOR SEQ ID NO:44:	•			,
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 42 base pairs			1	
(B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear			.*	
(5) 10102001. 1111011				÷.
(xi) SEQUENCE DESCRIPTION: SEQ ID	NO:44:		÷	
TGAAACAAAG CACTATTGCA CTGGCACTCT TACC	GTTACC GT			4:
(2) INFORMATION FOR SEQ ID NO:45:				
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 42 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear		•	*) * *	
(2), 20202001, 2011020			•	
(xi) SEQUENCE DESCRIPTION: SEQ ID	NO:45:			
TACTGTTTAC CCCTGTGACA AAAGCCGCCC AGGTC	CAGCT ĞC			42
(2) INFORMATION FOR SEQ ID NO:46:		* .	;	*
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 44 base pairs (B) TYPE: nucleic acid				
(C) STRANDEDNESS: single (D) TOPOLOGY: linear	•	•	:	
(x1) SEQUENCE DESCRIPTION: SEQ ID	10:46:		*	
TCGAGTCAGG CCTATTGTGC CCAGGGATTG TACTAC	GTGGA TCCG		•	44
(2) INFORMATION FOR SEQ ID NO:47:	0			
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 38 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear		*	,	
			* .	

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 42 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ. ID NO:47:

TGGCGAAAGG GAATTCGGAT CCACTAGTAC AATCCCTG

	NCE DESCRIPTION: SEQ		* .		42
GGCACAATAG GCC	TGACTCG AGCAGCTGGA CC	CAGGGCGGC TT	8		42
	N FOR SEQ ID NO:49:				-
(A) (B) (G)	NCE CHARACTERISTICS: LENGTH: 42 base pairs TYPE: nucleic acid STRANDEDNESS: single TOPOLOGY: linear		× · · · · ·	le	1 1
131	we ÷	¥.	· .		
(xi) SEQUE	NCE DESCRIPTION: SEQ	ID NO:49:	9	***	
TTGTCACAGG GGT	AAACAGT AACGGTAACG GT	AAGTGTGC CA	*		42
*	* * * * * * * * * * * * * * * * * * * *		,		
(2) INFORMATION	N FOR SEQ ID NO:50:				
(A) (B) (C)	NCE CHARACTERISTICS: LENGTH: 42 base pairs TYPE: nucleic acid STRANDEDNESS: single TOPOLOGY: linear		10	, , , , , , , , , , , , , , , , , , ,	
					*
(-1) SPONE	NCE DESCRIPTION: SEQ	ID NO:50:			
The state of the s	TIGITIC ACTITATITI CI				42
					÷ '
	N FOR SEQ ID NO:51:			. *	7
(A) (B) (C)	NCE CHARACTERISTICS: LENGTH: 21 base pair: TYPE: nucleic acid STRANDEDNESS: single TOPOLOGY: linear	s			
<u> </u>					
(x1) SEQUE	ENCE DESCRIPTION: SEQ	ID NO:51:			
TAACGGTAAG AG	IGCCAGTG C	*			
(52) INFORMAT	ION FOR SEQ ID NO:52:	1.0		9	
(A) (B)	ENCE CHARACTERISTICS: LENGTH: 68 base pair TYPE: nucleic acid STRANDEDNESS: single TOPOLOGY: linear	LS			
(ix) FEAT (A) (B) (D)	TURE:) NAME/KEY: misc_diff) LOCATION: replace(2) OTHER INFORMATION: MIXTURE OF A AND LOCATIONS 28, 31	/note= "M REPRI			

	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:52:		
AGCT	CCCG	GA TGCCTCAGAA GATGMNNMNN MNNMNNMNNM	NNMNNMNNMN NGGCTTTTGC	. 60
	\GGGG			68
		RMATION FOR SEQ ID NO:53:		
	(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 54 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear		
- n*	(ix)	FEATURE: (A) NAME/KEY: misc_difference (B) LOCATION: replace(17, "") (D) OTHER INFORMATION: /note= "M RE MIXTURE OF A AND C AT THIS I LOCATIONS 20, 23, 26, 29, 32		
,	· (v:1)	SEQUENCE DESCRIPTION: SEQ ID NO:53:		٠.,
	-	A TCCGCCMNNM NNMNNMNNMN NMNNMNNMNN M	NNMNNATGM GAAT	. 54
•		MATION FOR SEQ ID NO:54:		
		SEQUENCE CHARACTERISTICS: (A) LENGTH: 27 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear		
•				
((xi)	SEQUENCE DESCRIPTION: SEQ ID NO:54:		
GGTAA	ACAG	T AACGGTAAGA GTGCCAG		27
(2) I	NFOR	MATION FOR SEQ ID NO:55:		
, a	(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 19 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear		
((xi)	SEQUENCE DESCRIPTION: SEQ ID NO:55:		
		GC CACAGGGGT	***	19
		RMATION FOR SEQ ID NO:56:		
		SEQUENCE CHARACTERISTICS: (A) LENGTH: 63 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: lin ar		· · · · · · · · · · · · · · · · · · ·

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:	60
(x1) SEQUENCE DESCRIPTION AGGGTCATCG CCTTCAGCTC CGGATCCCTC AGAAGTCATA AACCCCCCAT AGGCTTTTGC	63
CAC	
(2) INFORMATION FOR SEQ ID NO:57:	1.
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 47 base pairs (A) LENGTH: pucleic acid	
(C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:	47
TCGCCTTCAG CTCCCGGATG CCTCAGAAGC ATGAACCCCC CATAGGC	
(2) INFORMATION FOR SEQ ID NO:58:	•
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 25 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(B) 1070L061. 12110-1	7
(x1) SEQUENCE DESCRIPTION: SEQ ID NO:58:	
CAATTTTATC CTAAATCTTA CCAAC	25
(2) INFORMATION FOR SEQ ID NO:59:	*
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 21 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	**
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:	21
GCCTTCAGCC TCGGATCCGC C	
(2) INFORMATION FOR SEQ ID NO:60:	*, *
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 21 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:	21
CGGATGCCTC AGAAGCCCCN N	7

- (2) INFORMATION FOR SEQ ID NO:61:
 - (i) SEQUENCE CHARACTERISTICS:

 (A) LENGTH: 30 base pairs

 (B) TYPE: nucleic acid

 (C) STRANDEDNESS: single

 (D) TOPOLOGY: linear
 - (x1) SEQUENCE DESCRIPTION: SEQ ID NO:61:

CGGATGCCTC AGAAGGGCTT TTGCCACAGG

30

I CLAIM:

- 1. A composition of matter comprising a plurality of cells containing a diverse population of expressible oligonucleotides operationally linked to expression elements, said expressible oligonucleotides having a desirable bias of random codon sequences produced from random combinations of first and second oligonucleotide precursor populations having a desirable bias of random codon sequences.
 - 2. The composition of claim 1, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is unbiased.
 - 3. The composition of claim 1, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is biased toward a predetermined sequence.
 - 4. The composition of claim 1, wherein said first and second oligonucleotides having random codon sequences have at least one specified codon at a predetermined position.
 - 5. The composition of claim 1, wherein said cells are procaryotes.
 - 6. The composition of claim 1, wherein said cells are E. coli.

- for the expression of a diverse population of random peptides from combined first and second oligonucleotides having a desirable bias of random codon sequences, comprising: two vectors: a first vector having a cloning site for said first oligonucleotides and a pair of restriction sites for operationally combining first oligonucleotides with second oligonucleotides; and a second vector having a cloning site for said second oligonucleotides and a pair of restriction sites complementary to those on said first vector, one or both vectors containing expression elements capable of being operationally linked to said combined first and second oligonucleotides.
 - 8. The kit of claim 7, wherein said vectors are in a filamentous bacteriophage.
 - 9. The kit of claim 8, wherein said filamentous bacteriophage are M13.
 - 10. The kit of claim 7, wherein said vectors are plasmids.
 - 11. The kit of claim 7, wherein said vectors are phagemids.
 - 12. The kit of claim 7, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is unbiased.
 - 13. The kit of claim 7, wherein the desirable bias of random codon sequences of said first and second oligonucl otides is diverse but biased toward a predetermined sequence.

- 14. The kit of claim 7, wherein said first and second oligonucleotides having a desirable bias of random codon sequences have at least one specified codon at a predetermined position.
- 15. The kit of claim 7, wherein said pair of restriction sites are Fok I.
- peptides from diverse populations of combined first and second oligonucleotides having a desirable bias of random codon sequences, comprising: a set of first vectors having a diverse population of first oligonucleotides having a desirable bias of random codon sequences and a set of second vectors having a diverse population of second oligonucleotides having a desirable bias of random codon sequences, said first and second vectors each having a pair of restriction sites so as to allow the operational combination of first and second oligonucleotides into a contiguous oligonucleotide having a desirable bias of random codon sequences.
 - 17. The cloning system of claim 16, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is unbiased.
 - 18. The cloning system of claim 16, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is diverse but biased toward a predetermined sequence.
 - 19. The cloning system of claim 16, wherein said first and second oligonucleotides having a desirable bias of random codon sequ nces have at least one specified codon at a predetermined position.

. . . .

- 20. The cloning system of claim 16, wherein said combined first and second vectors is through a pair of restriction sites.
- 21. The cloning system of claim 16, wherein said pair of restriction sites are Fok I.
- 22. A composition of matter comprising a plurality of cells containing a diverse population of expressible oligonucleotides operationally linked to expression elements, said expressible oligonucleotides having a desirable bias of random codon sequences.
 - 23. The composition of claim 22, wherein said cells are procaryotes.
 - 24. The composition of claim 22, wherein said expressible oligonucleotides are expressed as peptide fusion proteins on the surface of a filamentous bacteriophage.
 - 25. The composition of claim 22, wherein said filamentous bacteriophage is M13.
 - 26. The composition of claim 22, wherein said fusion protein contains the product of gene VIII.
 - 27. The composition of claim 22, wherein said diverse population of oligonucleotides having a desirable bias of random codon sequences are produced from the combination of diverse populations of first and second oligonucleotides having a desirable bias of random codon sequences.

- 28. The composition of claim 22, wherein the desirable bias of random codon sequences of said oligonucleotides is unbiased.
- 29. The composition of claim 22, wherein the desirable bias of random codon sequences of said oligonucleotides is diverse but biased toward a predetermined sequence.
- 30. The composition of claim 22, wherein said oligonucleotides having a desirable bias of random codon sequences have at least one specified codon at a predetermined position.
- 31. A plurality of vectors containing a diverse population of expressible oligonucleotides having a desirable bias of random codon sequences.
- 32. The vectors of claim 31, wherein said oligonucleotides are expressible as fusion proteins on the surface of filamentous bacteriophage.
- 33. The vectors of claim 31, wherein said filamentous bacteriophage is M13.
- 34. The vectors of claim 31, wherein said fusion protein contains the product of gene VIII.
- 35. The vectors of claim 31, wherein the desirable bias of random codon sequences of said oligonucleotides is unbiased.
- 36. The vectors of claim 31, wherein the desirable bias of random codon sequences of said oligonucl otides is diverse but biased toward a predetermined sequence.

- 37. The vectors of claim 31, wherein said oligonucleotides having a desirable bias of random codon sequences have at least one specified codon at a predetermined position.
- 38. A composition of matter, comprising a diverse population of oligonucleotides having a desirable bias of random codon sequences produced from random combinations of two or more oligonucleotide precursor populations having a desirable bias of random codon sequences.
- 39. A method of constructing a diverse population of vectors having combined first and second oligonucleotides having a desirable bias of random codon sequences capable of expressing said combined oligonucleotides as random peptides, comprising the steps of:
 - (a) operationally linking sequences from a diverse population of first oligonucleotides having a desirable bias of random codon sequences to a first vector;
 - (b) operationally linking sequences from a diverse population of second oligonucleotides having a desirable bias of random codon sequences to a second vector; and
 - (c) combining the vector products of steps (a) and (b) under conditions where said populations of first and second oligonucleotides are joined together into a population of combined vectors capable of being expressed.

10

- 40. The method of claim 39, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is unbiased.
- 41. The method of claim 39, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is diverse but biased toward a predetermined sequence.
- 42. The method of claim 39, wherein said first and second oligonucleotides having a desirable bias of random codon sequences have at least one specified codon at a predetermined position.
- 43. The method of claim 38, wherein steps (a) through (c) are repeated two or more times.

10

15

- 44. A method of selecting a peptide capable of being bound by a ligand binding protein from a population of random peptides, comprising:
 - (a) operationally linking a diverse population of first oligonucleotides having a desirable bias of random codon sequences to a first vector;
 - (b) operationally linking a diverse population of second oligonucleotides having a desirable bias of random codon sequences to a second vector;
 - (c) combining the vector products of steps (a) and (b) under conditions where said populations of first and second oligonucleotides are joined together into a population of combined vectors;
 - (d) introducing said population of combined vectors into a compatible host under conditions sufficient for expressing said population of random peptides; and
 - (e) determining the peptide which binds to said ligand binding protein.
- 45. The method of claim 44, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is unbiased.
- 46. The method of claim 44, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is diverse but biased toward a predetermined sequence.

- 47. The method of claim 44, wherein said first and second oligonucleotides having a desirable bias of random codon sequences have at least one specified codon at a predetermined position.
- 48. The method of claim 44, wherein steps (a) through (c) are repeated two or more times.

20

- 49. A method for determining the nucleic acid sequence encoding a peptide capable of being bound by a ligand binding protein which is selected from a population of random peptides, comprising:
- of first oligonucleotides having a desirable bias of random codon sequences to a first vector;
- (b) operationally linking a diverse population
 of second oligonucleotides having a
 desirable bias of random codon sequences
 to a second vector;
 - (c) combining the vector products of steps (a) and (b) under conditions where said populations of first and second oligonucleotides are joined together into a population of combined vectors;
 - (d) introducing said population of combined vectors into a compatible host under conditions sufficient for expressing said population of random peptides;
 - (e) determining the peptide which binds to said ligand binding protein;
 - (f) isolating the nucleic acid encoding said peptide; and
 - (g) sequencing said nucleic acid.

- 50. The method of claim 49, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is unbiased.
- 51. The method of claim 49, wherein the desirable bias of random codon sequences of said first and second oligonucleotides is diverse but biased toward a predetermined sequence.
- 52. The method of claim 49, wherein said first and second oligonucleotides having a desirable bias of random codon sequences have at least one specified codon at a predetermined position.
- 53. The method of claim 49, wherein steps (a) through (c) are repeated two or more times.
- 54. A method of constructing a diverse population of vectors containing expressible oligonucleotides having a desirable bias of random codon sequences, comprising operationally linking a diverse population of oligonucleotides having a desirable bias of random codon sequences to expression elements.
- 55. The method of claim 54, wherein said oligonucleotides are expressible as fusion proteins on the surface of filamentous bacteriophage.
- 56. The method of claim 54, wherein said filamentous bacteriophage are M13.
- 57. The method of claim 54, wherein said fusion protein contains the product of gene VIII.

- 58. The method of claim 54, wherein the desirable bias of random codon sequences of said oligonucleotides is unbiased.
- 59. The method of claim 54, wherein the desirable bias of random codon sequences of said oligonucleotides is diverse but biased toward a predetermined sequence.
- 60. The method of claim 54, wherein said oligonucleotides having a desirable bias of random codon sequences have at least one specified codon at a predetermined position.
- 61. The method of claim 54, wherein said operationally linking further comprising the steps of:
 - (a) operationally linking a diverse population of first oligonucleotides having a desirable bias of random codon sequences to a first vector;
 - (b) operationally linking a diverse population of second oligonucleotides having a desirable bias of random codon sequences to a second vector; and
 - (c) combining the vector products of steps (a) and (b) under conditions where said populations of first and second oligonucleotides are joined together into a population of combined vectors.
- 62. The method of claim 61, wherein steps (a) through (c) are repeated two or more times.

- 63. A method of selecting a peptide capable of being bound by a binding protein from a population of random peptides, comprising:
 - (a) operationally linking a diverse population
 of oligonucleotides having a desirable
 bias of random codon sequences to
 expression elements;
 - (b) introducing said population of vectors into a compatible host under conditions sufficient for expressing said population of random peptides; and
 - (c) determining the peptide which binds to said ligand binding protein.
- 64. The method of claim 63, wherein said population of random peptides are expressed as fusion proteins on the surface of filamentous bacteriophage.
- 65. The method of claim 63, wherein said filamentous bacteriophage are M13.
- 66. The method of claim 63, wherein said fusion protein contains the product of gene VIII.
- 67. The method of claim 63, wherein the desirable bias of random codon sequences of said oligonucleotides is unbiased.
- 68. The method of claim 63, wherein the desirable bias of random codon sequences of said oligonucleotides is diverse but biased toward a predetermined sequence.

10

- 69. The method of claim 63, wherein said oligonucleotides having a desirable bias of random codon sequences have at least one specified codon at a predetermined position.
- 70. The method of claim 63, wherein step (a) further comprises:
 - (a1) operationally linking a diverse population of first oligonucleotides having a desirable bias of random codon sequences to a first vector;
 - (a2) operationally linking a diverse population of second oligonucleotides having a desirable bias of random codon sequences to a second vector; and
 - (a3) combining the vector products of steps (a) and (b) under conditions where said populations of first and second oligonucleotides are joined together into a population of combined vectors.
- 71. The method of claim 70, wherein steps (al) through (a3) are repeated two or more times.

- 72. A method of determining the nucleic acid sequence encoding a peptide capable of being bound by a ligand binding protein which is selected from a population of random peptides, comprising:
 - (a) operationally linking a diverse population of oligonucleotides having a desirable bias of random codon sequences to expression elements.
- (b) introducing said population of vectors

 into a compatible host under conditions

 sufficient for expressing said population

 of random peptides;
 - (c) determining the peptide which binds to said ligand binding protein;
 - (d) isolating the nucleic acid encoding said peptide; and
 - (e) sequencing said nucleic acid.
 - 73. The method of claim 72, wherein said population of random peptides are expressed as fusion proteins on the surface of filamentous bacteriophage.
 - 74. The method of claim 72, wherein said filamentous bacteriophage are M13.
 - 75. The method of claim 72, wherein said fusion protein contains the product of gene VIII.
 - 76. The method of claim 72, wherein the desirable bias of random codon sequ nc s of said oligonucleotides is unbiased.

10

- 77. The method of claim 72, wherein the desirable bias of random codon sequences of said oligonucleotides is diverse but biased toward a predetermined sequence.
- 78. The method of claim 72, wherein said oligonucleotides having a desirable bias of random codon sequences have at least one specified codon at a predetermined position.
- 79. The method of claim 72, wherein step (a) further comprises:
 - (a1) operationally linking a diverse population of first oligonucleotides having a desirable bias of random codon sequences to a first vector;
 - (a2) operationally linking a diverse population of second oligonucleotides having a desirable bias of random codon sequences to a second vector; and
 - (a3) combining the vector products of steps (a) and (b) under conditions where said populations of first and second oligonucleotides are joined together into a population of combined vectors.
- 80. The method of claim 78, wherein steps (al) through (a3) are repeated two or more times.
- 81. A vector comprising two copies of a gene encoding a filamentous bacteriophage coat protein, both copi s encoding substantially the same amino acid sequence but having different nucleotide sequences.

- 82. The vector of claim 81, wherein said filamentous bacteriophage is M13.
- 83. The vector of claim 81, wherein said gene is gene VIII.
- 84. The vector of claim 81, wherein said vector has substantially the sequence shown in Figure 5 (SEQ ID NO: 1).
- 85. A vector comprising two copies of a gene encoding a filamentous bacteriophage coat protein, one copy of said gene capable of being operationally linked to an oligonucleotide wherein said oligonucleotide can be expressed as a fusion protein on the surface of said filamentous bacteriophage or as a soluble peptide.
- 86. The vector of claim 84, wherein said one copy of said gene is expressed on the surface of said filamentous bacteriophage.
- 87. The vector of claim 84, wherein said bacteriophage coat protein is M13 gene VIII.

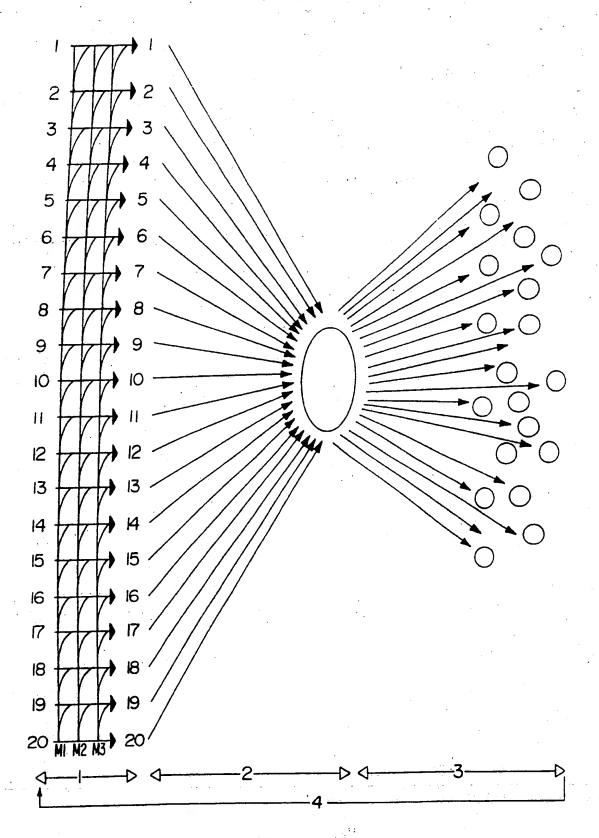
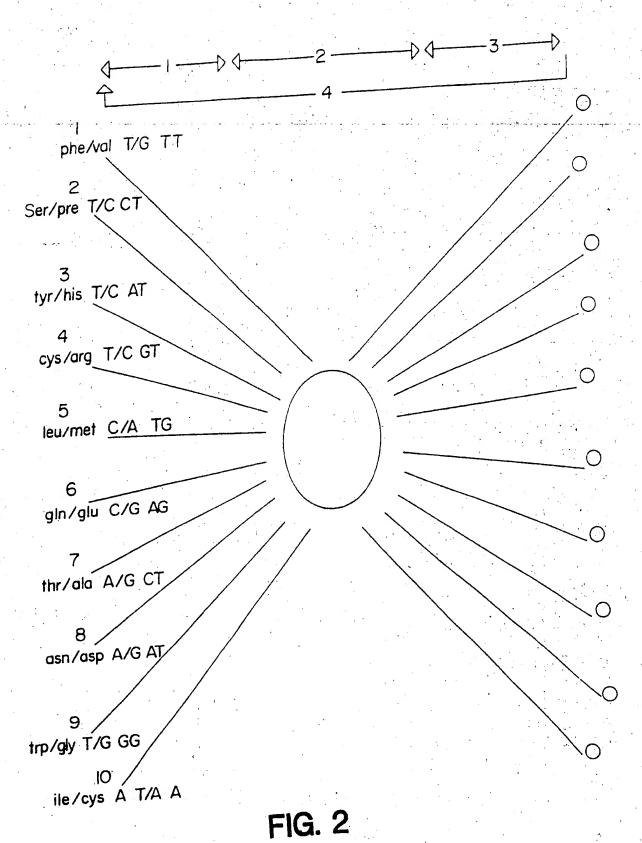
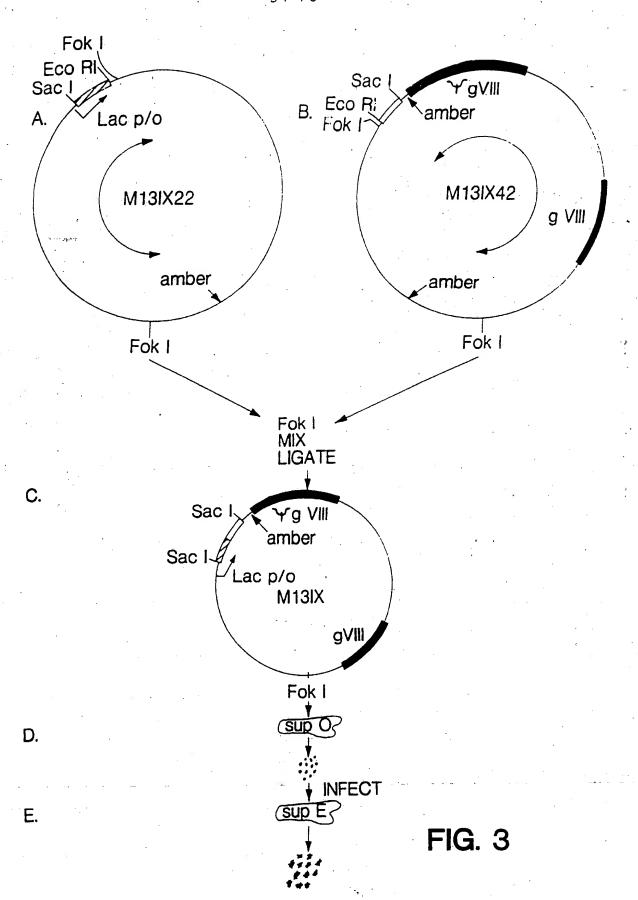


FIG. 1

SUBSTITUTE SHEET





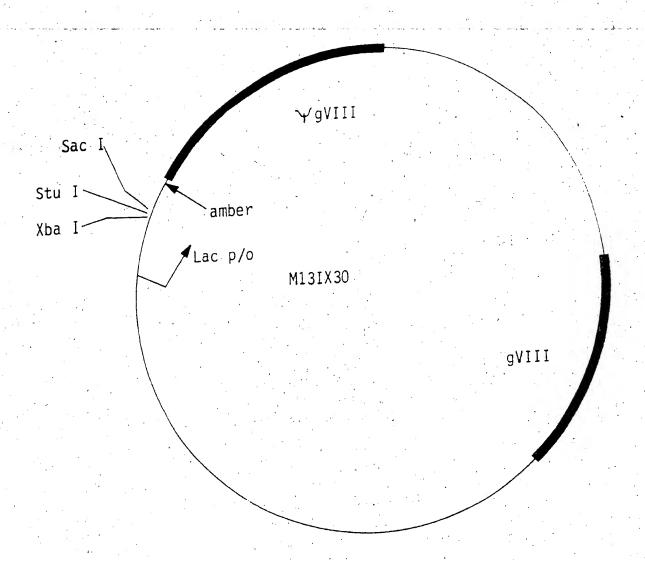


FIG. 4

SUBSTITUTE SHEET

FIG. 5-1

```
TTATCACACG GTCGGTATTT CAAACCATTA
                                                                                         AACGCCTTAT
GCTTACTAAA
ATCAGCATTT
TCAGACCTAT
                                                   ATTCTTATTT
                                                                                                                                                                  ATATAACCCA ACCTAAGCCG
AATTCACTAT TGACTCTTCT
               TCCGGTGTTT
                                                    AGAĂGATGAA
                                                                                                                              ACATATAGTT
                AATTTAGGTC
                                                    TTGGATTTGC
AGGTAGTCTC
ATCTAAGCTA
TACAGAAGCA
TAATTCAAAT
3901
                TGTCTTGCGA
                                                                                                                              GATTTTGATA
                                                                                                                                                                                                        ATTAATTAAT
3961
                                                                                                                                                                  CTAAGGGAAA
                                                                                                                                                                                                      TACTGTTTCC
CTTGATGTTT
GCGCGATTTT
                GAGGTTAAAA
                                                                                                                              TTCAAGGATT
4021
                                                                                          TCGCTATGTT
                                                                                                                                                                  TTGATTTATG
                                                                                         AGGTTATTCA
GAAATTGTTA
TCAGGTAATT
ATCAGGCGAA
                                                                                                                              CTCACATATA.
                CAGCGTCTTA
                                                                                                                                                                                                                                            4260
                                                                                                                                                                   ATTTTGTTTT
                AGCGACGATT
                                                                                                                              AATGTAATTA
4141
                                                                                                                                                                  ATTCGCCTCT
TTTCTCCCGA
                                                                                                                              GAAATGAATA
                                                                                                                                                                                                       TGTAAAAGGT
                 ATTAAAAAAGG
                                                                                                                                                                                                                                            4380
4201
                GTTTCATCAT
GTAACTTGGT
                                                      CTTCTTTTGC
                                                                                                                              TCCGTTATTG
                                                                                                                                                                                                       CTTTATTTCT
TTAGAAGTAT
                                                      ATTCAAAGCA
                                                                                                                                                                   TACGCAATTT
                                                                                                                             CCTGAAAATC
GGTTCAATTC
                                                                                                                                                                                                                                            4500
                                                                                           TGACGTTAAA
                                                                                                                                                                  CTTCCATTAT
                                                      TATATTCATC
                ACTGTTACTG
GTTTTACGTG
AATCCAAACA
GATAATTCCG
                                                                                                                                                                                                                                            4560
                                                                                                                                                                                                        GGAATATGAT
                                                                                           TGATATGGTT
4381
                                                      CTAATAATTT
                                                                                                                              TTGCCATCAT
GTTCCGCAAA
                                                                                                                                                                                                        TACTCAAACT
                                                                                           TATTGATGAA
                                                                                                                                                                  <u>ĂŢĞAŢAAŢĞŢ</u>
                                                     ATCAGGATTA
CTCCTTCTGG
ATAACGTTCG
CTAAATCCTC
4441
                                                                                                                                                                 TTGTCGAATCT
GCTCTAATCT
TTTCTACTGT
AGCAAGGTGA
GCGGTTAAGCAAGGTGAA
                                                                                          TGGTTTCTTT
                                                                                                                                                                                                        GTTTGTAAAG.
4501
                                                                                                                              TTAATACGAG
TCTATTGACG
CCTCAATTCC
TTTGAGGTTC
                                                                                          GGCAAAGGAT
AAATGTATTA
AGATAACCTT
GGGTTTGATA
                                                                                                                                                                                                        ATTAGTTGTT
4561
                 TTTAAAATTA
                                                                                                                                                                                                        TGATTTGCCA
4621
                TCTAATACTT
                                                                                                                                                                                                        TĞCTTTĂĞĂT
4681
                                                                                                                                                                                                                                            4860
                                                      AAGATATTTT
                AGTGCACCTA
ACTGACCAGA
                                                                                                                                                                                                       TACTGACCGC
CGATGTTTTA
TGTGCCACGT
CCCTTTTATT
4741
                                                       TATTGATTGA
                                                                                         TCAGCGTGGC
TGCTGGTGGT
AAAGACTAAT
GAAGGGTTCT
ATCTGCCTATT
GCGATAGTTTC
                                                                                                                               ACTGTTGCAG
TCGTTCGGTA
4801
                                                      CTGCTGGCTC
TTTTATCTTC
                                                                                                                                                                   TTTTTAATGG
                 TTTTCATTTG
                                                                                                                             TCGTTCGGTA
AGCCATTCAA
ATCTCTGTTGCAA
AGTTCTTCTAA
AGTTCTTCTAA
AATTTGCGTG
CGCTCTGATT
GCCCTGTAGC
ACTTGCCAGC
CGCCCGGCTTT
TTTAGGGCAC
GCCTGATAG
CTTGTTCCAA
GATTTTGCCG
                                                                                                                                                                 AAATATTGTC
GCCAGGAATGAC
TGGCTGGCAAG
CTCAGGCAAG
ATGGACAGGTA
GCGTACGAGGA
4861
4921
4981
                                                                                                                                                                                                                                            5040
                 GGGCTATCAG
ATTCTTACGC
ACTGGTCGTG
                                                       TTCGCGCATT
                                                                                                                                                                                                       GATTGAGCGT
TAATATTGTT
                                                     TTTCAGGTCA
TGACTGGTGA
GTATTTCCAT
 5041
                                                                                                                                                                                                       TGATGTTATT
TCTTTTACTC
CCTGTCTAAA
AAGCACGTTA
5101
                  CAAAATGTAG
                                                                                          CGATAGTTTG
TACAACGGTT
AAACACTTCT
GTTTAGCTCC
CATAGTCCCC
  5161
                                                      CCAGCAAGGC
GAAGTATTGC
                  CTGGATATTA
                 ACTAATCAAA
GGTGGCCTCA
ATCCCTTTAA
                                                                                                                                                                                                                                            5460
                                                       CTGATTATAA
TCGGCCTCCT
                                                                                                                                                                    CCAACGAGGA
                                                                                                                                                                                                        GCĞČĞĞĞĞĞĞ
ÇÇĞCTÇÇTTT
                                                                                                                                                                    GGCGCATTAA
GCCCTAGCGC
                                                                                                                                                                                                                                             5580
  5401
                                                        TCAAAGCAAC
                   TACGTGCTCG
TGTGGTGGTT
                                                      ACGCGCAGCG
CCTTCCTTTC
TTAGGGTTCC
GGTTCACGTA
ACGTTCTTTA
                                                                                            TGACCGCTAC
                                                                                                                                                                                                         CTCTAAATCG
                                                                                                                                                                    CCCCGTCAAG
CTCGACCCCA
                                                                                            ŢĊĠĊĊĂĊĠŢŢ
                                                                                                                                                                                                         AAAAACTTGA
                                                                                                                                                                                                                                              5700
                   CGCTTTCTTC
                                                                                           GATTTAGTGC
GTGGGCCATC
ATAGTGGACT
                                                                                                                                                                                                                                              5760
   5581
                                                                                                                                                                                                         GCCCTTTGAC
                                                                                                                                                                  ACGGTTTTTC
ACTGGAACAA
                                                                                                                                                                                                        CĂCTCĂĂČCĆ
                                                                                          ATAGTGGACT

ATTTATAAGG

GCAAACCAGC

GCAAACCAGC

GCTGTTGCCC

GCTGTTGCCC

GCTGTTGCCC

AAGCGGCAG

CTCTCCCCGC

AAGCGGCAG

CTTTACACTT

TATGCTTCCC

ACACAGGAAA

CTAGGCTGAA

CTAGGCTGAA

GCCCGCACCG

TGAGCTATGAC

TGAGTACATT

TAAATTATTC

AAAAAAGCTTT

TAAATTATTC

TGAGTTCCGG

AACGTACCGT

TCACATTTAA

ATTTTAACAA

AATTTTAACAA

AATTTTAACAA

AATTTTAACAA

AATATTCATCGA

TGATTTCATCACC

ATATTCATCGA

ATTTTCATCGA

ATTTTCATCGA

ATTTTCATCGA

ATTTTCATCGA

ATTTTCATCGA

TGATTTCACCT

ATTTCATCGA

ATTTTCATCGA

TGATTTCACCT

ATTTCATCGA

ATTTTCATCGA

TGATTTCACCT

ATTTCATCGA

ATTTTCATCGA

ATTTTCATCGA

TGATTTCACCT

ATTTCATCGA

TGATTTCACCT

ATTTCATCGA

TGATTTCACCT

ATTTCATCGA

TCACCT

TCACCT
   5641
                                                                                                                                                                                                        CACCATCAAA
CTCTCAGGGC
AACCACCCTG
GCAGCTGGCA
                   TTTGGGTGAT
GTTGGAGTCC
   5701
                                                                                                                                                                    ATTTCGGAAC
TGCTGCAACT
                                                                                                                                                                                                                                              5940
                   TATCTCGGGC
CAGGATTTTC
CAGGCGGTGA
GCGCCCAATA
CGACAGGTTT
                                                        TATTCTTTTG
                                                                                                                                                                                                                                              6000
    5821
                                                                                                                                                                    TGAAAAGAAA
                                                         GCCTGCTGGG
                                                                                                                                                                                                                                             6060.
                                                        AGGGCAATCA
CGCAAACCGC
CCCGACTGGA
GCACCCCAGG
TAACAACTTC
    5881
                                                                                                                                                                     ATTCATTAAT
                                                                                                                                                                                                          TGAGTTAGCT
    5941
                                                                                                                                                                     GCAATTAATG
                                                                                                                                                                    GCTCGTATGT
CAGGATGTAC
CTGCTAAGGC
GGGCTATGGT
CGAGCAAGGC
                                                                                                                                                                                                          ŢĢŢĢŢĢĢĀĀŢ
    6001
                                                                                                                                                                                                          GAATTCGCAG
TGCATTCAAT
    6061
                     CACTCATTAG
                                                                                                                                                                                                                                               6300
                                                                                                                                                                                                          AĞTAĞTTATA 6360
TTCTTAACCA 6420
CGCAĞCCTĞA 6480
                     TGTGAGCGGA
GTAGGAGAGC
                                                          TCGGCGGATC
CAAGTGCTAC
     6241
                      AGTTTACAGG
     6301
                                                                                                                                                                       CCAACGATTG
GGTGCCGGAA
                                                           CCATAGGGAT
                      GTTGGTGCTA
                                                                                                                                                                                                            AGCTGGCTGG 6540
ATGCACGGTT 6600
     6361
                     GCTGGCGTAA
ATGGCGAATG
AGTGCGATCT
ACGATGCGCC
CCACGGGGGAAA
                                                           TAGCGAAGAG
                                                          GCGCTTTGCC
TCCTGAGGCC
CATCTACACC
TCCGACGGGT
GACGCGAATT
     6421
                                                                                                                                                                       AAACTGGCAG
GGTCAATCCG
TGTTGATGAA
                                                                                                                                                                                                           CCGTTTGTTC
AGCTGGCTAC
ATGAGCTGAT
AAATATTTGC
ATATGATTGAT
TCTCAGGCAA
                                                                                                                                   ATCCCATTAC GGTCAATCCG CCGTTTGTTC
TCACATTTAA TGTTGATGAA AGCTGGCTAC
GCGTTCCTAT TGGTTAAAAA ATGAGCTGAT
AATATTAACG TTTACAATTT AAATATTTGC
CTGATTATCA ACCGGGGTAC ATATGATTGA
TTCTCTTGTT TGCTCCAGAC TCTCCAGGCAA
AATAGCTACC CTCTCCGGCA TTAATTTATC
TGATTTGACT GTCTCCGGCC TTTCTCTCACCC
CATTGCATTT AAAATATATG AGGGTTCTAA
CTCTCAGGCT TTATTGCTTA ATTTTGCTAA
CTCTGAGGCT TTATTGCTTA ATTTTGCTAA
CGTT
      6481
       6601
      6661
6721
6781
                        ĀĞĞĀĀĞĞCÇA
                                                             TTTAACGCGA
TTCCTGTTTT
TTACGATTAC
GCCTTTGTAG
GTTGAATATC
TTACCTACAC
                                                                                                                                                                                                                                                  6900
                         TTAACAAAAA
                                                                                                                                                                                                                                                   6960
                        TTATACAATC
                                                                                                                                                                                                                                                    7020
        6841
                        CATGCTAGTT
                                                                                                                                                                                                                                                    7080
        6901
        6961
7021
                                                                                                  ATATTGATGG
ATTACTCAGG
                         AGCTAGAACG
                                                                                                                                                                                                                                                     7200
7260
                         TTTTGAATCT
                                                                                                 AAATAAAGGC
TAGCTTTATG
ATTTATTGGA
30
          7081
                                                              CCTTGCGTTĞ
ACAACCGATT
TGCCTGTATG
20
                          AAATTTTAT
         7141
7201
7261
                          TGTTTTTTGGT
                          TTCTTTGCCT
```

FIG. 5-2

FIG. 6-1

		7000
		GTCGGTATTT CAAACCATTA 3900
	AACGCCTTAT TTATCACACG	OLYSSTATE TEGERAL OF TARREST
3841 TCCGGTGTTT ATTCTTATTT		TANAN YORK ACCTABGULG ASSES
3841 (19222) CCTC AGAGAIGAN	ALCACCATTT ACATATAG!!	DIDIDION SEXT TOACHELLE HUOU
3901 DOLLARYCOCK TIGGALLIOU	HILLOUGH AT CATTTIGALA	
SUPPLY THE TAXABLE AGGLAGICION	- LUDGER FORT TTCAAGGALL	CTAAGGGAAA ATTAATTAAT 4200
AUZI UNUSCHETTA ATTIANGUIT		
-1001 CAGUILLIA DISIRACO		AA11 [G11] G11384777 11320
LITE AGCGALUAL LOVETTONA	TGAAALIGIL AAALIYAAAA	AATT (GUUTU 1909 CARACC 11380
ASOL ATTAAAAAAG SISSATTTO	CTCAGGIAAI IUGGGAAAAT	GT1161666 #122122277776 ###0
LIGGI TOTTILLATUR LYTTCAAAACI	· AATLAGGUGA ALGESTATAT	CTACGUAALL ISSEARIARIA ASOO
TGTAAUITUU LALLAYPPOAT	CTGACGITAA AUSTYCAATT	
1551 TACTGITAUL DIGISTATE	r TTGATATUUT TUULLERATEA	· +0+0,000 000
TGTTTTALUT OUTSACCAT		A ATT A I A A I I I I I I I I I I I I I
		SPECTOCAAT TOTTTGIAAA 4000
	S CLOCKLACEN TITAALALUA	ULLUS PARTO TATTAGTIGE 4/40
4201 192122 AATAAUGIII	2 COUCESTATE ATOTATIOAL	TTGATTTGLL 4800
· 4021	! <u>とひつつよとして</u> です。すぐですぐみみしまし	SICCALCATE ATGCT LAGA 4999
4001 LISTOCACCT VOVENIALI	L ISSOFFFOAT ATTIGAGE	COCCACATA ATACIGALLO 4340
4/42 INSTANCAGE ATAILGAIL	S SESSOCIATE CACTULIOUA	COUNTY TO COCATGILL 4980
4801 222772777, 60161, 1660	1 6162585556 エエしばエコーロー	
4001 LLASICCTCT GTT LAILI	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	
4921 CLICATION GTTCGCGCA	1 AAAAGAGAGAGA TATOTOTGTT	GGCCAGAAIG SSSTACIOCO EIGO
1,001 Malalat FATOR CLASSICCT		CCAT LICAGA SUNTANTET 5220
		N ATGGC 16666 919717777 5280
Fin1 TAY (60 (CO) 912077777	A TGAGCGIIII 1964PAAAAA	r ACTCAGGCAA GIGAIPIACT EZAO
	G. CCGATAGITI SAGATERA	T GATGGALAGA LILLITATATA EMON
5221 TCTGGALALL ACARGTATT	G CTACAACGG! LATILLETC	T GGCGTACCOT TOURS ASSETT SUGO
5281 TACTAATCAA AGAAGTATAT	·A AAAACACIIO IQQQQQAAAX-	T TOCAACGAGO MAASSCCCCC ESOO
5341 CGGTGGCCTC ACTGGCCTC	TGTT IAGUTU CCCCAPAPA	c regregiation decayateatt reson
	VA CLUTAGIACO COSESCO	
	CTCKCILL LA LACTIOUS	
EE21 GTGIGGIGGI LACTTOCT		A CCTCGACCCC AAAAAACTTG 5700
		A COLOCATATA CGCCCIII LON 2/00
Fein GGGGGLLLUU LLLTFERIC		
		A COPTECCIÓN CONCONTRA 2000
	^ . T T I K I K N IG I I I I I I I I I I I I I I I I I	L COLOTO TOTO LAGGO 2240
	II SOCIAL CONG. COTGGALLU	TO THE POUNT AND AND CAUCULE DUUG
5821 010781777 CCCCTGC	CO CONTRACTOR CETETICSCI	TGCAGCIGGC BUSY
DOOL NECKCOCCC AAGGGLAP	10 4055566666611660	CU GALLYTTAAT GTGAGIIAGU GIZO
DAAT COURSESSAAT ACCUAAAL	LU LEISTECT GTGAGLULI	AH COCCENTRE TIGITOLOGAA DIOC
6001 GOCCECTT TOCCGAU	GO AMPRESACACT TTATGCIII	CO COCIPEINAT ACCTALIBLE DATE
PODT TOUCHER TO GENERAL COLOUR	THE STATESTAN GGAGACAG	IL ALAMIACCCC ACCTCGTGAL 6000
DIZI TECECOCCE ATAACAA	ILL LAPPYOFFEE TECCCAAL	CA CECALARCE TANGETIGGE DOCUMENT
DIOI LINASCACCE GETHINAL	ICI ICANCCARTO AGTOTIAN	III CIRCOSTATOCO MACTIANICO 0429
PAGE INCOME TO CONTRACT	ILL CLESTON CTC CGAAAALL	CL OUGE SECT CLALING DAGO
6301 GACCCAGACT CLAGAA	MAL DICCIPCETA CCCTANID	IOU UMASSEFECT TTO GGUALU OJAO
ロンロエ ごろうちょうじょくしゃ しかいいいい	CIT LYSYSAATCE CGAAIGUL	JUL 1119886171 CERTIFICAL DOUG
6421 CCCTTCCCAA CAGTTGC	GCA GULLBANIES CENTETTO	COL GOSTACANCE TABLLIATED DOGS
6481 ((() 155545 566644	IDUI OUGESSYAAA TECELUI	AIC TOUCHER ACTOGULAL OVER
~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~		[[[]
6601 CULLICATTO ATTOCA	LUUI LISTIFATOON NGGCLAG	ACG CGAA! ALL TANCANATA 6840
6661 CATTACECTE GATGAA	AGCT GGCIACAGGA AGGAAT	TTA ACGCGAALLI LAMERTCHCA CONO
FFOI ATTIANION SOLVAN	メチェル・ドレービター・ドラン・クロウムさつ	TCC TG[1][[[000 0011217767 6060
6701 TOCIALIUUI LOOPPY	WAAT ATTIGULIAL HUNDIUS	TAC GATTALLUTT YOLYYYXXXXX 7070
6841 TTAALGILLA CCCTAC	ATAT GATIGALALO CTONTAC	FOOT TIGINGALUL GLYDDAYFAAT JORO
6901 TTATCAACCG GGGTAL	TCTC AGGCAATGAC CTGATAG	ETTE AATAILAIAI IUMIOSCATT 7140
6901 TTATCAACCG GGGAAAAAAAAAAAAAAAAAAAAAAAAA	TTAA TTTATCAGCT AGAALGI	TTAC CIALALALIA VIVEZZZZZZZ 7200
6961 CTTGTTTGCT CCGGC	TONGELLLE UNDIVI	COTT GOGILIGAANI MAAYYEEEEE 7060
7001 TTGALIGICI 20207	FACCE TICTARARAL LILLS+	COGATTTAGC THAIGCILL 7520
71/1 6001118888 19191	27566 4601001011 117227	TOCC TOTATGATTT ATTGGALGE 7320
7141 GCATTTAAAA TATTA 7201 CCCGCAAAAG TATTA	エスティイ・エルバー ひかしししし しょうししりょ	IGCC TOTAL EN
4664 CAGGCTTTAT IGUI	AATTT TGCTAATTCT TIGGOT	40
7261 GAGGO. 10	20	

FIG. 6-2

			.	•		•
			- 1 7	o ¹ 1 3		0 1 60
	1	0 12				
	1 AATGCTACT	A CTATTAGTA	G AATTGATGC	C ACCTTTTCA	G CTCGCGCCC	C AAATGAAAAT 60
_	I WALCOLVO!			A AATGTATCT		
6.	I ATAGCTAAA	C AGGTTATTG	A CCATTTGCG			
12		A ATTGGGAAT	C AACTGTTAC	A TGGAATGAA	A CTTCCAGAC	A CCGTACTTTA 180
			TOACCTACA			
18.	1 GTTGCATAT				C WOCHWIIW	
24.	I TCTGCAAAA	A TGACCTCTT	A TCAAAAGGA	G CAATTAAAG	G. TACTCTCTA	A TCCTGACCTG 300
57	TTCCACTTT			T GAAGCTCGA	A TTAAAACGC	G ATATTTGAAG 360
30.	1 TTGGAGTTT					
36.	I TETTTCGGG	Č ŤŤĊĊŤĊŤŤĂ <i>i</i>	A TCTTTTTGA	T GCAATCCGC	T TIGCTICIG	
42.				G TCATTCTCG	T TTTCTGAAC	T GTTTAAAGCA 480
	LAGOGIAAA					
483	L TTTGAGGGG	G ATTCAATGA/	A TATTTATGA	C GATTCCGCA	G TATTGGACG	
54				ACTICITITE	G CAAAAGCCT	C TCGCTATTTT 600
				TĂTĠĂŤÁĠŤŮ		TATGCCTCGT 660
601			ANACONOUG			
661	AATTCCTTTT	「 GGCGTTATG7	ATCTGCATTA	\ GTTGAATGT0	G GTATTCCTA	
7 21	ATGAATCTTT		TAATGTTGTT	CCGTTAGTTC	C GTTTTATTA <i>P</i>	CGTAGATTTT 780
781	TCTTCCCAAC	GTCCTGACTG	GTATAATGAG	CCAGTTCTTA		
841	CAATGATTAA	AGTTGAAATT	AAACCATCTC	AAGCCCAATT	TACTACTCGT	TCTGGTGTTT 900
			TCACTGAATG	AGCAGCTTTG		TTGGGTAATG 960 "
901	CTCGTCAGGG					
961	*AATATCCGGT	: TCTTGTCAAG	ATTACTCTTG	<u>ATGAAGGTCA</u>	GCCAGCCTAT	
1021	TGTACACCGT	TCATCTGTCC	TCTTTCAAAG	TTGGTCAGTT	CGGTTCCCTT	ATGATTGACC 1080
	CTCTCCCCCT	CGTTCCGGCT	AAGTAACATG	GAGCAGGTCG		CACAATTTAT 1140
1081	GTCTGCGCCT		ANGIANCATO	TOTTTCCCCC	TTCCTATAAT	
1141	CAGGCGATGA	TACAAATCTC	CGTTGTACTT	TGTTTCGCGC	TTGGTATAAT	CGCTGGGGGT 1200
1201	CAAAGATGAG	TGTTTTAGTG	TATTCTTTCG	CCTCTTTCGT	TTTAGGTTGG	TGCCTTCGTA 1260
1501		CTATTTTCC		ĂĂĂČŤŤĊČŤĊ		CTTTAGTCCT 1320
1261	GTGGCATTAC	GTATTTTACC			TOTTTCCCTC	
1321	CAAAGCCTCT	GTAGCCGTTG	CTACCCTCGT	TCCGATGCTG	TCTTTCGCTG	CTGAGGGTGA 13805
1381	CGATCCCGCA	AAAGCGGCCT	TTAACTCCCT	GCAAGCCTCA	GCGACCGAAT	ATATCGGTTA 1440;
	TOCCTCCCCC	ATCCTTCTTC	TCATTCTCCC		GGTATCAAGC	TGTTTAAGAA 1500%
1441	TGCGTGGGCG	ATGGTTGTTG		CGCAACTATC	GGIAICANGC	
1501	ATTCACCTCG	AAAGCAAGCT	GATAAACCGA	TACAATTAAA	GGCTCCTTTT	GGAGCCTTTT 1560
1561	TTTTTGGAGA	TTTTCAACGT	GAAAAAATTA	TTATTCGCAA	TTCCTTTAGT	TGTTCCTTTC 1620'
		CCCCTCAAAC				AGAAAATTCA 1680
1621	TATTCTCACT	CCGCTGAAAC	TGTTGAAAGT	TGTTTAGCAA	AACCCCATAC	
1681	TTTACTAACG	TCTGGAAAGA	CGACAAAACT	TTAGATCGTT	ACGCTAACTA	TGAGGGTTGT 1740
1771	CTCTCCAATC	CTACAGGCGT	TGTAGTTTGT	ACTGGTGACG	AAACTCAGTG	TTACGGTACA 1800
1741	CTGTGGAATG	CIACAGGCGI	TATCCCTCA	ACTUOTORCO	CTCCCTCTCA	GGGTGGCGGT 1860
1801	TGGGTTCCTA	TTGGGCTTGC	TATCCCTGAA	AATGAGGGTG	GTGGCTCTGA	
1861	TCTGAGGGTG	GCGGTTCTGA	GGGTGGCGGT	ACTAAACCTC	CTGAGTACGG	TGATACACCT 1920:
	ATTCCCCCCCT	ATACTTATAT	CAACCCTCTC		ATCCGCCTGG	TACTGAGCAA 1980
1921	ATTCCGGGCT	AIACTIATA	TTCTCTTCTC	CACTCTCACC	CTCTTAATAC	TTTCATGTTT 2040
1981	AACCCCGCTA	ATCCTAATCC	TTCTCTTGAG	GAGTCTCAGC	CTCTTAATAC	
2041	CAGAATAATA	GGTTCCGAAA	TAGGCAGGGG	GCATTAACTG	TTTATACGGG	CACTGTTACT 2100
	CAUANIANIA		AACTTATTAC		CTGTATCATC	AAAAGCCATG 2160%
2101	CAAGGCACTG	ACCCCGTTAA	HACITALIAC		TCC4TTCTCC	
2161	TATGACGCTT	ACTGGAACGG	TAAATTCAGA	GACTGCGCTT	TCCATTCTGG	
2221	GATCCATTCG	TTTGTGAATA	TCAAGGCCAA	TCGTCTGACC	TGCCTCAACC	TCCTGTCAAT 2280
2201	CCTCCCCCC	GCTCTGGTGG	TGGTTCTGGT	GGCGGCTCTG	AGGGTGGTGG	CTCTGAGGGT 2340
2281	GCTGGCGGCG			CCCCCTTCCC	CTCCTCCCTC	TGGTTCCGGT 2400
2341	GGCGGTTCTG	AGGGTGGCGG	CTCTGAGGGA	GGCGGTTCCG	GTGGTGGCTC	
2401	GATTTTGATT	ATGAAAAGAT	GGCAAACGCT	AATAAGGGGG		AAATGCCGAT 2460
	GAAAACGCGC	TACAGTCTGA	CGCTAAAGGC	AAACTTGATT	CTGTCGCTAC	TGATTACGGT 2520
2461	GHAMACGCGC	ATCATICION	TCCTCACCTT	TOCCCCCTTC		TGGTGCTACT 2580
2521	GCTGCTATCG	ATGGTTTCAT	TGGTGACGTT	TCCGGCCTTG	CTCACCCTCA	TAATTCACCT 2500
2581	GGTGATTTTG	CTGGCTCTAA	TICCCAAATG	GUICAAGICG		TAATTCACCT 2640
2641	TTAATGAATA	ATTTCCGTCA	ATATTTACCT	TCCCTCCCTC	AATCGGTTGA	ATGTCGCCCT 2700
	TTTCTCTTTTA	CCCCTCCTAA	ACCATATGAA	TTTTCTATTC	ATTGTGACAA	AATAAACTTA 2760
2701	TTTGTCTTTA		ACCA LA LUAA		TTATCTATCT	
2761	TTCCGTGGTG		TCTTTTATAT	GIIGCLACCI	TTATGTATGT	ATTITCTACE 2820
2821	TTTĞCTAAÇA	TACTGCGTAA	TAAGGAGTCT	TAATCATGCC	AGTTCTTTTG	GGTATTCCGT 2880
2021		TTTCCTCCCT	TTCCTTCTGG	TAACTTTGTT	CGGCTATCTG	CTTACTTTTC 2940
2881			ATACCTATTC	CTATTTCATT	CTTTCTTCCT	CTTATTATT6 3000
2941	TTAAAAAGGG	CTTCGGTAAG	ATAGCTATTG	CTATTTCATT	GTTTCTTGCT	
รักกำ.	CCCTTAACTC	AATTCTTGTG	GGTTATCTCT	CTGATATTAG	CGCTCAATTA	CCCTCTGACT 3060
2001.	TTOTTOACCC	TOTTOACTTA	ATTOTOCOCT	CTAATGCGCT	TCCCTGTTTT	TATGTTATTC 3120
3061	TTGTTCAGGG	TGTTCAGTTA	ATTCTCCCGT	CIAMIDUCCI	I C C C T T	
3121	TCTCTGTAAA.	GGCTGCTATI	TTCATTTTTG	ACGITAAACA	AAAAATCGTT	
3181	ATTGGGATAA	ĀĪĀĀĪĀTĀĊĊ	TGTTTATTTT	GTAACTGGCA	AATTAGGCTC	TGGAAAGACG 3240
7101	CTCCTTACCC	TTCCTAACAT		ATTGTAGCTG	GGTGCAAAAT	AGCAACTAAT 3300
3241		TTGGTAAGAT	TCAGGATAAA	ALIUINUCIU		
3301	CTTGATTTAA	GGCTTCAAAA	CCTCCCGCAA	6 C666A66	TCGCTAAAAC	GCCTCGCGTT 3360
2261	CTTACARTAC	CGGATAAGCC	TTCTATATCT	GATTTGCTTG	CTATTGGGCG	CGGTAATGAT 3420
3361	たいけんながないから	AAAATAAAAA	~666 67	CTTCTCCATC	AGTGCGGTAC	TTGGTTTAAT 3480
3421	ILLIALBAIG	AAAATAAAA		CCCATTANTO	ATTCCTTTCT	ACATGCTCGT 3540
3481	ACCCGTTCTT	GGAATGATAA	GUAAAGACAG	CCGATTALIG	ALIUUITICI	
3541	AAATTAGGAT		TTTTCTTGTT	CAGGACTTAT	CTATTGTTGA	TAAACAGGCG 3600
コンオナ	DON'T LOOK	TACCTCAACA	TETTETTÄ	TGTCGTCGTC	TGGACAGAAT	TACTTTACCT 3660
3601	CGTTCTGCAT	TAGCTGAACA	TGTTGTTTAT		TOCCTOTOCO	TAXATTACAT 2700
3661	TTTGTCGGTA	CTTTATATTC	TCTTATTACT	UULILGAAAA	TGCCTCTGCC	TAAATTACAT 3720
2771	CTTCCCCTTC	TTAAATATGĞ	CGATTCTCAA	TTAAGCCCTA	CTGTTGAGCG	TTGGCTTTAT 3780
J/ Z I	טווטטטטווט	1177717170	CONTROLONA	, ,	J. J J J J J J J	

FIG. 7-1

	0				TAATTATGAT 3840
		CCCATATGAT	ACTAAACAGG	CITTITICIAG	I D D I D I D I D I D I D I D I D I D I
3781 ACTGGTAAG	A ATTIGIATAA	AACGCCTTAT		GTCGGIALLI	ACCCCTTCTT 3960
3841 TCCGGTGTT	1 4116714577	GCTLAULAAA	- ATATATIIGA	AAAAGTTTTC ATATAACCCA	ACCTAAGCCG 4020
3841 TCCGGTGTT 3901 AATTTAGGT		ΔTCAGCALL	ACATATAGIT	AATTCACTAT	TGACILITE 4000
3961 16161166		TCAGACCIAI	GATTTTGATA	CTAAGGGAAA	ATTAALLAAL 4200
4021 GAGGTTAAA 4081 CAGCGTCTT	ΓΑ ΔΤΟΤΑΑΘΟΙΑ	TOGOTALGIL		TTGATTIALG	TACTGITIOU 4200
	TT TACAGAAGUA	AGGTTATTCA	AAATGTAATT	AATTTTGIII	TOTAL 750
1201 ATTAAAAA	IG GTAATILAAA	TGAAATTGTT	TGAAAIGAAI	AATTEGECTE	TĞCĞCĞATTT 4320 ATĞTAAAAĞĞ 4380
4261 TGTTTCATO	N 101161116		ATCCGLIAL	13 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	TCTTIALLIU 4440
4321 TGTAALLY	コロ ・「ひ!」とひひひとと	CTGACGILAA	ACCTGAAAAI	CITY COATAA	TTCAGAAGTA 4200
4381 AC16116		TTGATAIGGI	TGGTTCAATT	TOTGATABLE	AGGAATATGA 4200
4441 TGTTTTACO	AC AATCAGGALI	ATATTGATGA	TGTTCCGCAA	AATGALAALU	TIAL LUMBBOOK SEED
4501 TAATCCAA 4561 TGATAATT	on antentitie	GTGGTTTCTT GGGCAAAGGA	: TTTAATACGA	GTTGTCGAAL	1911 1710
1621 TTTTAAAA	サイトダイクト	CAAATGTATI	- ATCTATTGAU	GGCTCTAATC	TTGATT HILL 4000
4681 GTCTAALAG	~ [TAGATAALLI	TCCICAALIL	CTTTCTACTG	ATGCTTTAGA 4800
4741 TAGTGLAC	AG ATATTGALIG	AGGGTTTGAL	ATTTGAGGTT CACTGTTGCA	GGCGGTGTTA	ATACTGACCG 4920
4801 AACTGACCA	TT GCTGCTGGCI	CTCAGCGIGG	CACIBLICA	ATTTTTAATU	שבים הכל כל בחווח
4861 TITLLAT	CT GTTTIAIUII	CTGCTGGTGG TAAAGACTAA	. TAGCCALIUA	AAAATATIGI	L101227787 E100
AGRI AGGGCIAI	CA GIILGCGCC		TATCTCIGII	GGCCAGGG	CENTIGAGED DIOU
5041 TATTCLLA	re rilicassi	: AATCTGCCA <i>f</i>	TGTAAALAAL	CCATTTCAGA ATGGCTGGCG	GTAATAIIGI 2440
5101 AL. 16610		TGAGCGTTT			GTGATGTIAL 2200
5161 TCAAAATG 5221 TCTGGATA	TT ACCAGCAAG	LUGALAGI	GAGTTCTTCT TAATTTGCGT	GATGGACAGA	111.111.1001 27.22
5221 TCTGGATA 5281 TACTAATC	AA AGAAGTALIU	CINCHACOC	TCAAGAIIUI	GGCGTACLGI	1001010177 5/160
EZĂĂ CGGTGGCC	TC ACTGALLAL	1 サガサガラいろうし	· COGOTO I GAI	していればいってから	AGCGCGGCGG 2240
5401 AATCCCIL	TA ALCOGCOTO	A CCATAGIAC	G CGCCCIGIAS		CCCGCTCCT1 220V
5461 ATACGTGC		r GTGACCGCII	4 CACLIDECT	r TOCCCGIVA	1 00 10 17 CTTC 6700
5521 GTGTGGTG 5581 TCGCTTTC	TT CCCTTCCTL	I LILLELLACE		CCTCGACCCU	AAAAAAAAAAAA E760
EETT GGGGGCTU	CC TTTAGGGIL	L CGALLIAGE	T CGCCCIGATA	A GACGGIIII	ACACTCAACC 5820
5701 ATTTGGG	CH TOOLTHURE	T AATAGIGGA	C TCTTGTTCC	H HHCTOGAGO	A CCACCAILAA 200U
5761 CGIIGUAS	316 646677877	T GATTTALAA	G GGALLINGS	r TTGCTGCAA	C TCTCTCAGGE 2340
5821 CTATCTCO 5881 ACAGGAT	FFT CACCTGEIG	6 GOCHAHOCA	G CGTGGACCG C CGTCTCGCT	G GTGAAAAGA	A AAACCACC EDED
5881 ACAGGAI 5941 CCAGGCG	ETG AAGGGCAAI	C AGCIGINGS	c cacatteel	C GATICALIA	A LUCAUCHTOCK ETON
ENNI GGCGCCC	AA I ALGUMARES		À GTGAGCGCA	A - CGCAX 1+32	G TTGTGTGGGAA DLOU
6061 ACGALAG		is sctttacau	T TTATGULLY	C 00010010+	T TTACAALGIL 0240
6121 TCACTCA 6181 TTGTGAG	CEG ATAACAA!!	T CACACGLG	C ACTIONA	T GGAGAAAAI	A AAGTGAAALA 2360
	CCA AAACCCIG	10 01140000	TT ACCGLIACI	G TTTACCCC	A CTACTEGATE 6420
EZO1 AAGCACI	ATT GCACTGGL	AC ICTIACCO	~T ATTG1666	א טטטטטרוופי	G CAAGTGCTAC 6480
6361 [[5][[6]	יים במממים יום	rc ctgctaau	GC TGCALICA!	41 HOT 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	7 A CCN I MISISIS A L 00740
6421 CIAGGUI	ATT GGCTACGC	TT GGGCIAIG	01 40146115	ra atagugaa	GA GGCCCGCACC GGGO
6481 TGAGTAC	TTO AAAAAGII	TA CONSCRIP		AT GGCGCTTI	66 61941 66646 6720
GENT GATCGC	CCTT CCCAALAG	11 90000000	TR GAGTGCGA		AC CAACGTAACC 6780
FRET GCACCAU	GAAG COGIGES	ca GATGCAU	GT TACGAIGC		GG TTGTIALILE DOTO
6721 616616		rrc GCCG1114	III CCCHCGG		AT TATTIIIGAI DEUD
6781 TATCCC 6841 CTCACA	TTTA ATGTTGA	rga aagctggi	TA CAGGAAGO	INA ALITAALU	100 nn 7020
KANT GGCGTT	CCTA TTGGILA	HAM WAISTSH		AAT CTICKIG	HA COSTTOATES 7080
GOGT ANATAL	TAAL GILLAGE		TTG ACATGUI	AGI IIIACON	CTA GATCICILAA Z149
7021 TCTGAL	TAIL HACCOOS	AGA CTCTCAG	GCA ATGACCI	GAI AUCULTI	TAT CATALLUALU /200
7081 ATTCL	SCHOOL COTOTO	GGC ATTAATI	INI CARCINO	ATC TTTACCI	ACA CATTACICAS 7320
7141 AAATAC 7201 GTGAT	FIGAR TATCILL	GGC CTTTCIL	ALL VILLETT	TTA TCCIIGL	SIL PRACCETTAT 7380
7261 GCATIL	GCATT TAAAALA	ITAL CACCCT	*ATA 8761411	ו טט ואטתקטט	TAT CATTTATTEG /440
7323 (111.11	CCCGC ANAMOTA	STTTTAA TTTT	SCTA ATTETT	GCC TTGCCTG	/ 442
7381 GC <u>1C1</u>	DAUGU IIIAII			40	50 60
7441 ACGTT	10	20	30		
· · · · · · · · · · · · · · · · · · ·		the state of the s			

FIG. 7-2

FIG. 8-1

SUBSTITUTE SHEET

		the state of the s	8		AAACCATTA 3900
k.			CACACG	GTCGGTATTT C	444CC/ 66A
		ATTICTTATTT AACGCCTTAT	TTATCACACG	AAAACTTTTC . A	CCCC110117222
	3841 TCCGGTGTTT	A	ΑΤΔΙΔΙΙΙΟΝ	122127 X A C C C A A	CCT AAGUUG 4049
:			ACALALAULI	MICHARCE AT A	CALIFILLITIES
:			GATTLIBALA		I LANTINAL TERM
٠	3961 TGTCTTGCGA 4021 GAGGTTAAAA	AGGIAGIUIU LUGGATATOTT	TTCAAGGATI		AC 1511100 7400
	4021 GAGGGTCTTA	ATCIAAGUIA 1002777777	CTCACATATA		TTIGATULE 4200
,		TACAGAAGUA AGGILATTATT	AAATGTAATT		っこうじん ししょり コンピン・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・
	4141 AGCGACGATT	GTAALLCAAA LUQQQQAAXAA	TGAAATGAAT		TCTAAAA66 4200
	4201 ATTAAAAAAG			GTTTC/CCCC 5	CTTTATTTC 4440
	MOCI TALLILATUR	TATTCAAAGC AATCAGGCGA	ACCTGAAAAT	CTACGCAAL	TCAGAAGTA 4500
	ボラカェ TCTAAし ししじ	LANGE CTGACGINAM	TGGTTCAATT	CCTTCCATAA	11.80000
	7201 TACIGITACI	SAPARTATT TTGATALGOL	ATTOCCATO.	TOTGATABLE !	111111111111111111111111111111111111111
	ATAM TOTTLIALUI	ATATTGALGA	ATTGCCATCA	A STCALABLE	TAUTURATE SAGA
	JENI TAATULAAAU	COTOCTTOTA GTAGILIUI	THE ECOVITY	GTTGTUGAAT	
	JEGT TGATAATILL	COLLARDO GGGCAAAGGA	TTTAATACGA	GGCTCTAATC	HILINOITO: . COO .
	ICO1 TTTTAAAATI	L POLOGIATOR CANATGIALL	ATCTATTGAC	CTT I I I I I I I I I I	
	JEO1 GTC AA IAU	LIVICATATT TAGATAAUUI	TCCTCAATTC		TGCTTTAGA 4860
	ニュラルキ***Tがら けっしみししょ	TOUCHARTE ACCOUNTS	. Alliunuvi.	CCCCCCCTCTTA	STACTGALLG 4920
			I ALLIUI LOOM	OFFFFF AATG	SCGATGIIII 4300
			TTCGIILUUI	7 1 1 1 7 7 7 7 7 7	TGTGCCALG 2049
		r GTT HAILIL SIDERAGETA	TAGCCALIUM	CCCCACAATG	TCCCTITIAL 2100
	4921 CCTLACCTOR 4981 AGGGCTATCA		TATCICIGII	COLVETTONES	CGATTGAGUG 2100
	4981 AGGGCTATO	TTTILAGGIC AGARAGESSA	TGTAAATAAT	Y 2000 0 + 000 0 G	CTABIBLE DAAG
		r GTGALIGUIU AAISISPPP		HI GOOL GOOL	CTENIALIAI JEOU
	5101 TACIOGICA	A GGTATIILLA INAUENCTT		at it.Audoviii.	たてたて 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2
	5161 TCAAAATGT	T ACCAGLAAGO CECALACACA			TOT IGHT HAR DAOS
•	5221 TCTGGATAT	A AGAAGIALIO CIPCOTEFF		141111111111111111111111111111111111111	AAAGCACGTT 5460
	5281 186166667	C ACTGATIAIA AAAAAAAAAAA	r regetella	ししいれることから	ACCCCCGGGCGGCGGC
	5341 CGGTGGCCT	Λ. Δ 1.6600 00 ± 01 ± 1.75 = 1.0	E COCCCTGTAG		**************************************
	5401 AATCCCTTT		A CACTTGCCA	CGCCCTAGCG	CCTCTABALL JUNG
	5461 ATACGTGCT	C GILAMAGUAT GTGACCGCT			AAAAAACTTG 5700
		+ 6001100111 616066888		A COTOGALLUL	CGCCCTTTGA 5760
	Froi Trillill	L PPPACETTO CGALLIAGI		A GACGGIIII	
	5641 66666677			A AACTUGAACA.	
	2701 ATT 166610)	U 101147766	r GATTILLUAA	
	77C1. (C) 1(11AU)	IC COCCLAPTE CATTIALAR	ים מטבול ליבי	C TTGCTGCAAL	
•	FOOT FIAILILUS	コロ・ビスストナスとせたは、ほほじひひかししゃ	10 001000000	G GTGAAAAGAA	
	FOOT ACAINALL	LE POSSOCIATO AGOIDITO	\\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\	C GATTLATIAN	IN AUCTOC
	FAMI CCAMBILIO	IN TRACESTATIONS COTOLOGICAL	Lucui loca	A CGCAATIAAI	GIOTOTOTOTA 6180
	ZANA GGCGLLLA	AL MODERACTED ANAGERGE	A GIGAGGPP	C GGCTCGIAIG	
	ENGI ACGALAUU	II ICCOMPAND GOTTIALA	117100717	T GGCCGILUII	1125277776 6300
	čioi Tracillai	TA COCCACALACTE CACALACIA	IL MOTTOPES	IT GGAGAAAAI	4402674766 6360
	2101 TTGTGAGU	DO GLASSCATEGE GILALLUM	AU CITCHTA	rr rrtgibblaa	HANDER OF ENOU
	CONT GTGALIGO	ICH MANAGERACIA TOTTACCO	II WOLD TO		
	CZOL NAGCALIA	III GUNGLOS CCC ATOUGGGA	GC TOWNOOCC	AC GCTTGGGCT	
	6361 GGGGTTCA	TA CAGGCAAGTG CTACTGAG	TA CALLOCK	AC TITACGAGU	A AGGCTITITY SEAN
	CAOL CAALAGII	TTG CTTCTGAGGC ATCCGGGA TTA CAGGCAAGTG CTACTGAG TTA CAGGCAAGTG GGATTAAA		AC AGT HILLIGH	G CCTGAATTOC CEEN
	6481 TATAGTTO	101 991755CCCC [NCC1111]	.60 00118887	GC CGGAAAGCT	C CLIPPING OF GES
	Zeni AGCAATAU	BUB ADSOCTECT TOURGLAD	TON GRUDOSSI	CT GGCAGATGC	A CEELLACONI OCES
		GCT ILGCCLOPAC GGTCGTC		TCA ATCCGCCGT	T TELLILIAND UVO
	6661 GARCARU	LIG DUSSETTEET ARECTAL			C CCTALAGGAA DOTO
	čžoi GCGCLUA	さしし かいりいことがるそれ。 CTCGC1し	ACA TTTAAIG	110 01000000	IC I HAR ELIANO SEES
	6721 GCGCCCA 6781 GAGAATC	CGA CGGGTTGTTA CTCGCIC CGC GAATTATTT TGATGGC ACGC GAATTATTT AACAAAA	GTT CCTATTG		CA TTTGCLLAIR OZUU
		CGC GAATTATTTT TGALGGO	TAT TAACGTT	CGG GGTACATA	TO ATTIGACATOR 100
	COOL AAAAAII		MAIN TO THE PARTY OF THE PARTY	1.66 00172212	CVCCL DD 1 1940 C 7 202 1
	COCI CANILLI	TCCT GTT!!!!	TOTO TIGITIE	いしし していいいかん	AT TTAILAGUIN 1438
	7001 TAGUL	TACG ATTACCETC TOSSAS	ITAC CTALLL		CT CACCUITIU 4800
	7001 11-0140	ととてて てにしからなしししし しりつりりり	CATT TGACIL	1010 60055175	CT TOTAGARANT 1499
		TTGA ATATCATAL TCAGGC	ATTO CALLIA	AAAI DIDISKA	GT CATAAIGILL 4368
	7141 GAACGO 7201 AATCTT	THUE TOOMS A APA AAGIST	TOTO CCGUAR	AAU LLAGGE	
	7063 11141	CTTG COLLUMNS OF THITE	TCTG AGGCTT	TATT GCTTAAT	/403
	7201 AATCTT 7261 TITATC 7321 TTGGTA	CAAC CGATTTAGCT TTATGL		1	50 60
	7321 TTGGTA 7321 TGCCTT	GCCT GTATGATTTA TTGGAC	30	40	
	7381 TGCC 1	GCC1 GIATO 20	•		
	1		4		

FIG. 8-2

FIG. 9-1 SUBSTITUTE SHEET

		STOCCTATTT CAAACCATTA 3900
	• - • -	
	ATTETTATTT AACGCCTTAT	TIAILALACO OLOGATTITO ACCCGTILLE 3900
3841 TCCGGTGTTT		- A T A T A T T T LLAN. A A A A A C C C C C C C C C C C C C C
3841 10061814	- A C A A C A I (- B B) - INC I I M C I D D D	TARTYTICTT ATALANCULA ACCIDATES MOON
3901 AATTTAGGTC		- NOOTO - O - O - O - O - O - A A T T C N C F A L - HAAL FO F FO - O - O - O - O - O - O - O - O
		- CATTTTTCAID ABIICOUID : :====a+++AT /11/11
	. *************************************	**************************************
4021 GAGGTTAAAA	ATCTARGCTA TEGETATETT	TOURS AND THE ATTENDED TO THE TALL TO THE TABLE
4081 CAGCGTCTTA		TIMENTALA LACTOR CTTCATGILL 4200
TOOT ACCCAGGATT	TACAGAAGCA AGGTIALLYA	- 本本でである。いたないは、「いいは、」、「いく」、「こうなっては、「ない」、「いっ」。
4141 AGCGACGATT	*	- DD YEQQ 4 x - XTTCC(
JOON ATTAAAAAGG	LOUIST TONGGIAALI	GAAATGAATA QIIISEECCA TCTAAAAGGI 420U
4261 GTTTCATCAT		
	ATTENNISH A BILAGUCUM	
4321 GTAACTTGGT	TO TOUR LAND	CUIUMARAPAS ATTOCATTAT HAGAAUIAI 7208
4381 ACTGTTACTG		TOTAL TO SELECTIVE AND CONTRACT HODGE
	CTAATAATII 1941619616	
4441 GITITACOTO	ATCACCALIA LALIUNIUMO	- 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
4501 AATCCAAACA	ALCOSPICACE ACCULTUTION	611CCGCAAA 0199492477 CTTTGTAAA6:400U
		TTAATACGAG 119199796+ X++XGTTGTT 4/40
	- ATAACCIII (4 (515).AAAUUC)	- TOTALIGALGA GLIGITANIO
4621 TTTAAAATTA		- 1010112266 TTTTT181161, 1641110000, 1933
- λ691⊞TELAATAUTI	' Y!\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\	COLUMNITUDE INTERCACE TECTT TAGAI 4000
	- TATTCATIGA INGILLIUMIT	THE PROPERTY OF THE PARTY OF TH
4801 ACTGACCAGA	LIBERTON TO A COUNTIES	ACTUITUDE PARTITANTES (GATGIIIIA 4300
4861 TTTTCATTIG		
	- 	· ACCCATTCAA AAAIAIIUIU 1952-77877 6100
4921 CTCACCTCTG		こりしょうこうきょうし しししんじんい ロジュー・アンド・コード・カン・カーラー
HORT GEGELATUAU	1 1 1 2 2 1 2 2 7 2 2 7 C A CARCIGITIES	ALLICIO DOSCOSTA CATTGAGGG DIO
	1110400100 0000000000000000000000000000	・ とそんんんてんかしい じゅうしししかひかい シグミニミディとディッピングロ
	C TO ACTOCTOS ALLIBUUMMI	VICETAGE A TOCOTOGOGO TANTALIVAL ESSO
5101 ACTGGTCGTG	1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	CUIDITUDES AFRACCOMAG TGATGITATI 2200
5161 CAAAATGTAG	· [1] [A] [] [[[[[[[[[[[[[[[. ACTTCTT/ IN [] AUGUNDO YEAF-ACTC EXIII
	1	E AATTTCCCTG AIGGALAUNC LOLATOTAAA ENDO
	CARCIALIS IALAACGUI	DOI: 1122570 COCTACTO I LULIDIO 1000 SUCO
5281 ACTAATCAAA	* VOOSAPRATA AAACACIII	P CAAGAIICIO GGGIGSTECA AACCACGIIA DAGO
5341 GGTGGCCTCA	A LIUMINOUS CONTRACTOR	- CCCTCTCA11 1 AALGAOOD 002222222CCCC
5341 GGTGGCCTCTC	A TEST III I III I INDUCTO	E SCCCTCTAGC GGCGCALIAA GCGCGCGCTT ESRO
5401 ATCCCTTTA	Z ÷čižiččao CATAGIALU	~ GCCC G GCC
FART TACGTGUILL	LAMAGOCACCO TOACCOCTA	ALIIGULAGE COSSITERA CONTRADICO DOS
5521 TGTGGTGGT	ALUCUCACO LOGICA COT	+ cccccccll
	c ccttcciiii ilbillacoi	
5581 CGCTTTCTT	Y ==:::::::::::::::::::::::::::::::::::	L IIIACOUCHO ACCOTTITTO GOOGLIIGAL 2/00
5641 GGGGCTCCC	I I AUGUST CO COLLEGIO A	C GCCCTGATAG ACCOLLIGIA CACTCAACCC 5820
	T GGTTCACGTA GTGGGCCAT	* FFFCTTCCAA ALIGAACAA COYLYPSIII EOQO
	& ACCITCATTA ATABILBUAL	LULIULIUS ATTTCCCAAC CALLAIDAAN 2006
5761 GTTGGAGTC	C DUCTATTO ATTTALANG	G GAILLIUCCE TACATACT CTCTCAGGGC 2340
5821 TATCTCGGG	1 11 11 11 11 11 11 11 11 11 11 11 11 1	C GTGGACCGCI IQCIQCATAL XACCACCCTG 6000
	C GCCTGCIGGG GCMMMCCM	c ctctccclick lyananonon lyttagtccci coci
5881 CAGGALLL	A ACCCCABILA GUIDITOU	C DISTAGGGG ATTCATTAAL GUAGUIGGG STAA
5941 CAGGC6616	IN TOUGHT COOK CTOTOTOTO	I GUIT HOUSE AT TOACT AGUE DIZU
	D LUCHANCOU VISINA	* ************************************
	ccccx/	P +XTCCTTCCG GCTCGTATGT TOTOLSCCAC 62/10
6061 CGACAGGTT	LA SEXECCENCE CTTIACACI	I INITEDITION CAGATETAL GAALLUUVAS SEAS
- 6121 CACILATIA	NO CONCERNATION ACACAGGAS	A FAMILIATURE CONSCILLAR TOCKTIERAL DOUG
6181 TGTGAGCGG	A INNUMBILLY SELECTOR	(* CCCCATGA(): 1.1GL1880000 12272441748CZG():
	TURGUGGAIL LONGULIVE	P CCCTACCCTT GGGCTAIGUI AUIMUIICA CAOO
6241 GTAGGAGAG		I DOO INSEET CONCONNECT IT IN I MACON SIZE
EZNI AGTILALAU	JU COUPLESSORT TAARTIQL	LL VVVVILLE CONSCIENCE CCCVCCLIPY PAOR
SEC. CTTCCTLL	IA II.AIAUUUN	LA LTCCCCCCCCCCCCCCAALACTOCCCCCCCCCCCCCCCCC
6361 61166164	AA TAGCGAAGAG GUUUUU	SO SICCICIAGO GGTGCCGGAA AGCIGGGGT ECON
6421 GCTGGCGT	25 266777CCC TGG11100	GG CACCAGAAGC GGTGCCGGCAG ATGCACGGTT 6600
CHOI ATGGGGAAA	IN COMPLEXABLE CATACGGT	re ligitition haddings after coefficial poor
6541 AGTGCGAT	CT T(() GAGGCC 001172993	CT ATCC: A LLB: 1913 LLBB LOVE
6541 AGTGCGAL		OI DIVERSALA TOTTONICAN MINITORNI SESSO
ARNI ACHAIGCO	ILL PUTSTOCKE TELLVEL	GL ILACATITAD 1861-786AAAA ATGAGLIGAT 0/00
	an II.I.IIACOOO : :	TE GESTICETAL 1001144444 GAATATTTEE 6840
6661 CCACGOAG	CA GACGCGAATI ALLILIYS	(1) UCC () YY : 22
6721 AGGAAGGC 6781 TTAACAAA	CA PAPANACACA ATTITANI	WAY COLORS AND CECCE TO BE ALKINDED TO SECO
ETOI TTAACAAA	NAA TITAACEEETT	TT CTGATTATCA ACCEGGGGAC TCTCAGGCAA 6960
6841 TTATACAA		160
	5 TTACLALIAL LILLUMI	WA LATACCTACE CTETECOGULA ILAMILIACE TORO
6901 CATGCTAG	LI COCTTECTAG ATETELL	AAA AATAGUTAGU GACAAAAAA TTTCTCALLU ZUOO
FORT TRACCIO	AIA ULULI IPITO ATATTOM	reg Tealligaci giviyaya kecettetaa /140
	ACE ETTERALATE AIMITUS	
	ニネナ イモルアドーハームー みししみしょし	AGG CATTGCATTI AAAATATTAC AGGGTCATAA 7200
7081 TTTTGAA	LUL COTTOCCTTG AAATAAA	GGC IICICCOSCO TTATTCCTTA ATTITUCIAN 1200
71/11 ΔΔΔ1111	TALLULI 1966 AT TACCTTT	AGG CATTGCAGCA AAAGTATTAC AGGGTCATAA 7200 GGC TTCTCCCGCA AAAGTATTAC ATTTTGCTAA 7260 ATG CTCTGAGGCT TTATTGCTTA ATTTTGCTAA 7294
4884 PPP **	CCT. ACARCEGOUS INSTALLED	NIO OLEE
7201 TGTTTT	UUI IIVIIIIAAAAAA ATTIAII	GGA CGTT 40 50 60
7261 TTCTTTG	ונין וטניוטותים	30 40
7202 110	10 20	

FIG. 9-2

| 10 | 20 | 30 | 40 | 50 | 60 |
1 AATGCTACTA CTATTAGTAG AATTGATGCC ACCTTTCAG CTCGCGCCCC AAATGAAAAT 61 ATAGCTAAAC AGGTTATTGA CCATTTGCGA AATGTATCTA ATGGTCAAAC TAAATCTACT 121 CGTTCGCAGA ATTGGGAATC AACTGTTACA TGGAATGAAA CTTCCAGACA CCGTACTTTA 181 GTTGCATATT TAAAACATGT TGAGCTACAG CACCAGATTC AGCAATTAAG CTCTAAGCCA 181 TCTGCAAAAA TGACCTCTTA TCAAAAAGGAG CAATTAAAGG TACTCTCTAA TCCTGACCTG 181 TCTGCAAAAA TGACCTCTTA TCAAAAAGGAG CAATTAAAGG TACTCTCTAA TCCTGACCTG 181 TCTTCGGGC TTCCCTCTAA TCTTTTTGAT GCAATCCGCT TTCGTTCTGA CTATAATAGT 181 TCTTCGGGC TTCCCTCTAA TCTTTTTGAT GCAATCCGCT TTCCGTTCTGA CTATAATAGT 181 TCTGAGGGGG ATTCAATGAA TATTTATGAC GATTCCGCAG TATTGGACGC TATCCAGTCT 50 1 AAACCATTTA CTATTACCCC CTCTGGCAAA ACTTCTTTT CAAAAAGGCCC TACCCAGTCT 50 1 AATTCCTTTT GGCGTTATGT ATCTGCATTA GTTGAATGTG GTATTCCTAA ATCTCAACTG 7 1 ATGAATCTTT CTACCTGTAA TAATGTTGTT CCGTTAGTTC GTTTTATAA CGTAGATTTT 70 1 ATGAATCTTT CTACCTGTAA TAATGTTGTT CCGTTAGTT TACTACTCGT TCTGGTGTTT 90 1 CTCGTCAGG CAAGCCCTTAT TCTGGTGTTT 90 1 CTCGTCAGG CAAGCCCTTAT TCTGGTGTTT 90 1 CTCGTCAGG CAAGCCCTTAT TCTGGTGTTT 90 1 CTCGTCAGG CAAGCCTTAT TCTGGTGTTT 90 1 CTCGTCAGG CAAGCCTTAT TCTGGTGTTT 90 1 CTCGTCAGG CAAGCCTTAT TCACTGAATT AAACCATCTC AAGCCCAATT TACTACTCGT TCTGGTGTTT 90 1 CTCGTCAGG CAAGCCTTAT TCACTGAATT AAACCATCTC AAGCCCCAATT TACTACTCGT TCTGGTGTTT 90 1 CTCGTCAGG CAAGCCTTAT TCACTGAATT TACTACTCGT TCTGGTGTTT 90 1 CTCGTCAGG CAAGCCTTAT TCACTGAATT TACTACTCGT TCTGGTGAATG 90 1 CTCGTCAGG CAAGCCTTAT TCACTGAATG AGCAGCTTTG TTACCGTTGAT TTGGGTAATG 90 1 CTCGTCAGG CAAGCCTTAT TCACTGAATG AGCAGCTTTG TTACCACTGT TCTGGTTAATG 90 1 CTCGTCAGATG AGCAGCTTTG TTACCACTGT TCTGGTTAATGAG CCCCAATT TACTACTCGT TCTGGTTAATGAG CCCCAATT TACTACTCGT TCTGGTTAATGAG CCCCCAATT TACTACTCGT TCTGGTTAATGAG CCCCCAATT TACTACTCGT TCTGGGTAATG 90 1 CTCGTCAGATG AGCACCTTTAT TCACTACTCGT TCTGGTTAATGAG CCCCAATT TACTACTCGT TCTGGTTAATGAG CCCCAATT TACTACTCGT TCTGGTTAATGAG CCCCCAATT TACTACTCGT TCTGGTTAATGAG CCCCAATT TACTACTCGT TCTGGTTAATGAG CCCCAATT TACTACTCGT TCTGACTTAATGAG CCCCCAATT TACTACACATCATAATGAG CCCCAATT TACTACTCACATAATGAG CCCCAATT TACTACACACA 20 121 CGTTCGCAGA 180 181 GTTGCATATT TCAAAAGGAG
GGTTCGCTTT GAAGCTCGAA
TCTTTTTGAT GCAATCCGCT
TGATTTATGG TCATTCTCGT
TATTTATGAC GATTCCGCAG
CTCTGGCAAA ACTTCTTTTG
AAACGAGGGT TATGATAGTG
ATCTGCATTA GTTGAATGTG
TAATGTTGTT CCGTTAGTTC
GTATAATGAG CCAGTTCTTA
AAACCATCTC AAGCCCAATT
TCACTGAATG AGCAGCTTTG
ATTACTCTTG ATGAAGGTCA
TCTTTCAAAG TTGGTCAGTT
AAGTAACATG GAGCAGGTCG
CGTTGTACTT TGTTTCGCGC
TATCTTTCG CCTCTTCGT
CGTTTAATGG CAAACCTCCC
CTACCCTCGT TCCGATGCTG
TCATTGTCGG CGCAACTATC
GATAAACCGA TACAATTAAA
GAAAAAATTA TTATTCGCAA
TGTTGAAAACT TTAGATCGTT
TGTTAGTTGT ACTGGTGACG **30**0 301 360 600 601 GGTTTTTATC 661 AATTCCTTTT 721 ATGAATCTTT 781 TCTTCCCAAC 841 CAATGATTAA 901 CTCGTCAGGG 961 AATACCCGGT 660 720 GTATICCTAA ATCICAACTG 720
GTTTTATTAA CGTAGATTTT 780
AAATCGCATA AGGTAATTCA 840
TACTACTCGT TCTGGTGTTT 900
TTACGTTGAT TTGGGTAATG 960
GCCAGCCTAT GCGCCTGGTC 1020
CGGTTCCCTT ATGATTGACC 1080
CGGATTTCGA CACAATTTAT 1140
TTGGTATAAT CGCTGGGGGT 1200
ATGAAAAAGT CTTTAGTCTT 1320 CAAGCCTTAT TCTTGTCAAG TCATCTGTCC CGTTCCGGCT TACAAATCTC TGTTTTAGTG GTATTTTACC TGTACACCGT GTCTGCGCCT CAGGCGATGA 1021 CAAAGATGAG GTGGCATTAC CAAAGCCTCT CGATCCCGCA ATGAAAAAGT TCTTTCGCTG GCGACCGAAT CTTTAGTCCT GTAGCCGTTG AAAGCGGCCT CTGAGGGTGA 1380 1381 ATATCGGTTA 1440 TGCGTGGGCG ATTCACCTCG TTTTTGGAGA TATTCTCACT ATGGTTGTTG AAAGCAAGCT TTTTCAACGT CCGCTGAAAC GGTATCAAGC GGCTCCTTTT TTCCTTTAGT AACCCCATAC TGTTTAAGAA 1500 GGAGCCTTTT 1560 TGTTCCTTTC 1620 AGAAAATTCA 1680 AACCCCATAC ACGCTAACTA AAACTCAGTG GTGGCTCTGA CTGAGTACGG ATCCGCCTGG CTCTTAATAC TTTATACGGG CTGTATCATC TCCATTCTGG TGCCTCAACC AGGGTGGTGG 1680 1740 TTTACTAACG CTGTGGAATG TGGGTTCCTA TCTGAGGGTG ATTCCGGGCT AACCCCGCTA CAAGGCACTA TGAGGGTTGT
TTACGGTTACA
GGGTGGCGGT
TGATACACCT
TACTGAGCAA
TTTCATGTTT
CACTGTTACT
AAAAGCCATG 1681 TCTGGAAAGA ACTGGTGACG AATGAGGGTG ACTAAACCTC GACGGCACTT GAGTCTCAGC GCATTAACTG CAGTACACTC GACTGCGCTT TCGTCTGACC GGCGGCTCTG GGCGGTTCCG AATAAGGGGG CTACAGGCGT TTGGGCTTGC GCGGTTCTGA 1741 TGTAGTTTGT
TATCCCTGAA
GGGTGGCGGT
CAACCCTCTC
TTCTCTTGAG
TAGGCAGGGG
AACTTATTAC
TAAATTCAGA
TCAAGGCCAA
TGGTTCTGGT
CTCTGAGGGA TGTAGTTTGT 1860 1920 ATACTTATAT ATCCTAATCC GGTTCCGAAA ACCCCGTTAA 1980 2040 2100 2160 2041 CAAGGCACTG TATGACGCTT GATCCATTCG 2101 CTTTAATGAA TCCTGTCAAT CTCTGAGGGT TGGTTCCGGT ACTGGAACGG TTTGTGAATA **2220 228**0 GCTCTGGTGG AGGGTGGTGG GTGGTGGCTC **GCTGGCGGCG** AGGGTGGCGG ATGAAAAGAT TACAGTCTGA ATGGTTTCAT CTGGCTCTAA ATTTCCGTCA GGCGGTTCTG CTCTGAGGGA

FIG. 10-1

SUBSTITUTE SHEET

```
ATTOTTATTT AACGCCTTAT TTATCACACG GTCGGTATTT CAAACCATTA
                                                                                                                                                                        3960
                                                              GCTTACTAAA ATATATTTGA
ATCAGCATTT ACATATAGTT
TCAGACCTAT GATTTTGATA
TCGCTATGTT TTCAAGGATT
                                                                                                                                              <u>AČČŤĂAĠČĊĠ</u>
                                                                                                                                                                        4020
         TCCGGTGTTT
                                                                                                                    ATATAACCCA
                                     AGAAGATGAA
                                                                                                                                               TGACTCTTCT
                                                                                                                                                                        4080
          AATTTAGGTC
                                     TTGGATTTGC
AGGTAGTCTC
ATCTAAGCTA
                                                                                                                    AATTCACTAT
          TGTCTTĞCGĂ
                                                                                                                                               ATTAATTAAT
                                                                                                                     CTAAGGGAAA
TTGATTTATG
                                                                                        TTCAAGGALL
CTCACATATA
CTCACATATA
AATTTTGTTT
AAATGTAATT AATTTCGCCTC
ATCACATATA
ACCTGAAAAT
CTACGCAATT
ACCTGAAAAT
TGGTTCAATT
CCTTCCATAA
ATGGTAATC
ATTGCCATCA
AATGATAATG
                                                                                                                                              TACTGTTTCC
          GAGGTTAAAA
                                                               AGGTTATTCA
TGAAATTGTT
CTCAGGTAAT
AATCAGGCGA
CTGACGTTAA
TTGATATGCT
          CAGCGTCTTA
                                                                                                                                               TCTTGATGTT
                                      TACAGAAGCA
                                                                                                                                               TGCGCGATTT
           AGCGACGATT
                                      GTAATTCAAA
           ATTAAAAAAG
TGTTTCATCA
TGTAACTTGG
                                                                                                                                                ATGTAAAAGG
TCTTTATTTC
                                      TCTTCTTTIG.
                                                                                                                                                                         4440
                                      TĂTTCĂĂĂGC
GTATATTCĂT
                                                                                                                                                TTCAGAAGTA 4500
           TACTGTTACT
                                                                                                                                                AGGAATATGA
                                                                 TTGATATGGT
4381
                                      GCTAATAATT
                                                                                           ATTGCCATCA
TGTTCCGCAA
           TGTTTTACGT
TAATCCAAAC
TGATAATTCC
                                                                ATATTGATGA
GTGGTTTCTT
                                                                                                                     AATGATAATG
                                                                                                                                                TTACTCAAAC
                                      AATCAGGATT
GCTCCTTCTG
AATAACGTTC
TCTAAATCCT
                                                                                          TGTTCCGCAA
TTTAATACGA
ATCTATTGAC
TCCTCAATTC
ATTTGAGGTT
CACTGTTCGA
TTCGTTCGGT
TAGCCATTCA
TATCTCTGTT
TGTAAATAAT
TCCTGTTGCA
TAGTTCTCT
                                                                                                                     GTTGTCGAAT
GGCTCTAATC
CTTTCTACTG
CAGCAAGGTG
GGCGGTTAATA
                                                                                                                                                                         4680
                                                                                                                                                TGTTTGTAAA
4501
                                                                                                                                                TATTAGTTGT
                                                                GGGCAAAGGA
CAAATGTATT
                                                                                                                                               TTGATTTGCC
ATGCTTTAGA
ATACTGACCG
GCGATGTTTT
                                                                                                                                                                         4800
                                                                TAGATAACCT
AGGGTTTGAT
CTCAGCGTGG
CTGCTGGTGG
            GTCTAATACT
                                                                                                                                                                         4860
                                      AAAGATATTT
            TAGTGCACCT
                                      ATATTGATTG
GCTGCTGGCT
            AAČTĞĀCČĀĠ
                                                                                                                                                                          4980
           TTTTTCATTT
CCTCACCTCT
AGGGCTATCA
TATTCTTACG
TACTGGTTGTA
TCAAAATATT
                                                                                                                      ATTTTTAATG
                                                                                                                                                CTGTGCCACG
TCCCTTTTAT
                                                                                                                                                                          5040
                                                                                                                      AAAATATTGT
GGCCAGAATG
                                       GTTTTATCTT
                                                                 TAAAGACTAA
AGAAGGGTTC
                                                                                                                                                                          5100
                                       ĞTTCĞCĞČAT
CTTTCAĞĞTÇ
                                                                                                                                                CGATTGAGCG
                                                                                                                      CCATTTCAGA
ATGGCTGGCG
ACTCAGGCAA
                                                                AGAAGGGIIC
AATCTGCCAA
TGAGCGTTTT
CCGATAGTTT
CTACAACGGT
AAAACACTTC
TGTTTAGCTC
CCATAGTACG
GTGACCGCTA
CTCGCCACGT
CGATTTAGTG
 4981
                                                                                                                                                                          5220
                                                                                                                                                GTAATATTGT
                                       GTGACTGGTG
GGTATTTCCA
ACCAGCAAGG
 5041
                                                                                                                                                GTGATGTTAT
 5101
                                                                                                                                                CTCTTTTACT
TCCTGTCTAA
AAAGCACGTT
                                                                                           TAATTTGCGT
TCAAGATTCT
CCGCTCTGAAG
CACTTGCCAG
CACTTGCCAG
TCTCACCGCCA
                                                                                                                      GĂTĞGĂCĂGĂ
              TCTGGATATT
                                                                                                                                                                          5400
            TACTAATCAA
CGGTGGCCTC
AATCCCTTTA
ATACGTGCTC
GTGTGGTGGT
                                                                                                                      GGCGTACCGT
                                        AGAAGTATTG
                                                                                                                                                                           5460
                                        ACTGATTATA
                                                                                                                       TCCAACGAGG
                                                                                                                                                 AGCGCGGCGG
                                                                                                                      CGGCGCATTA
CGCCCTAGCG
TCCCCGTCAA
CCTCGACCCC
                                        ATCGGCCTCC
                                                                                                                                                 CCCGCTCCTT 5580
GCTCTAAATC 5640
                                                                 GTCAAAGCAA
  5401
                                        TACGCGCAGC
CCCTTCCTTT
TTTAGGGTTC
              TCGCTTTCTT
GGGGGCTCCC
                                        TGGTTCACGT
CACGTTCTTT
CTATTCTTTT
CGCCTGCTGG
              ATTTGGGTGA
   5701
              CGTTGGAGTC
   5761
               CTATCTCGGG
                                         AAGGGCAATC
ACGCAAACCG
              CCAGGCGGTG
GGCGCCCAAT
                                        TCCCGACTGG
GGCACCCCAG
ATAACAATTT
AAACCCTGGC
GCACTGCAC
               ACGACAGGTT
TCACTCATTA
               TTGTGAGCGG
GTGACTGGGA
                AAGCACTATT
    6301
                                          GGAGCTGAAG
                GAGGCATCCG
                                          GAGTACATTG
AAATTATTCA
ATCGCCCTTC
CACCAGAAGC
TCGTCCCCTC
ATCCCATTAC
                AAGTGCTACT
CATAGGGATT
GCCCGCACCG
                                                                     GGTGCCGGAA AGCTGGCTGG
AAACTGGCAG ATGCACGGTT
GGTCAATCCG CCGTTTGTTC
TGTTGATGAA AGCTGGCTAC
TGGTTAAAAA ATGAGCTGAT
TTTACAATTT AAATATTTGC
ACCGGGGTAC ATATGATTGA
TGCTCCAGAC TCTCAGGCAA
CTCTCCGGCA TTAATTTATC
GTCTCCGGCC TTTCTCACCC
AAAATATATG AGGGTTCTAA
AAAGTATTAC AGGGTCATAA
TTATTGCTTA ATTTTGCTAA
                 ŤĞĞŤŤŤCCĞĞ.
                 ĠĂŤĄĊĠĞŤČĞ
                                                                                                                          AGGAAGGCCA GACGCGAAT I TTAACAAAAA TTTAACGCGA TTTAACAATC TTCCTGTTTT CATGCTAGTT TTACGATTAC TGACCTGATA GCCTTTGTAG AGCTAGAACG GTTGAATATC TTTGAATCT TTACCTACAC AAATTTTAT CCTTGCGTTG TGTTTTGATTACT ACAACCGATT
                                           TCACATTTAA
GCGTTCCTAT
AATATTAACG
                  TGTTACTCGC
ATTTTTGATG
                  ATTITAACAA
                                             CTGATTATCA
TTCTCTTGTT
                                                                                                                                                                                7080
                ATÁTTGATÁG
ATÁTTGATGG TGATTGACT
ATTACTCAGG CATTGATT
ATTACTCAGG CATTGCATTT
AAATAAAGGC TTCTCCCGCA
TAGCTTTATG CTCTGAGGCT
ATTTATTGGA CGTT
       7021
                                                                                                                            TGTTTTGGT ACAACCGATT 7320
TTCTTTGCCT TGCCTGTATG 7380
        7201
         7261
                                                                                                                               | 50
                                                                                                                      40
                                                                         | 30
```

FIG. 10-2

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US91/07141

I. CLAS	SIFICATIO	N OF SUBJECT MATTER (if several class	sification symbols apply, indicate all)	
IPC(og to Internal 5): C12	ional Patent Cierrification (IPC) or to both N 1/24, 15/00; CO7H 21/00 35/252.33, 320.1, 172.3;	ational Classification and IPC 536/27	
	S SEARCI		the second se	
11 71220	J. J. LANG.		entation Searched 7	
Classificat	ion System		Classification Symbols	
Ciassincal	ion System			
u.s.		435/252.33, 320.1, 1	72.3, 69.1; 536/27	
		Documentation Searched other to the Extent that such Documents	than Minimum Documentation s are Included in the Fields Searched 8	
APS,	CAS: se	arch terms: Codon bins, o	odon preference	
III. DOCU	MENTS C	DNSIDERED TO BE RELEVANT		
Category *	Citatio	in of Document, 11 with indication, where app	ropriate, of the relevant passaged 12	Relevant to Claim No. 13
Y		A. 0.383.620 (Cook) 2 entire document.	2 August 1990,	1-87
Υ.	US 03 Ju	A, 4,458.066 (Caruthe uly 1984, see entire	rs et al.) document.	1-87
Y	US. 3 13 Se	4.771.000 (Verrips eptember 1988, see en	et al.) tire document.	8.9.24-26 32-34. 55-57.
				64-66 73-75.
Y	Volumet all of sy polypessen	ED MICROBIOLOGY AND E e 21. issued 1985. J, "Construction and nthetic DNA fragments eptides with elevated tial amino acids". pantire document.	M. Jaynes expression coding for levels of	3-87
"E" agriculture document of the control of the cont	ument defini sidered to bi independent g date ument which ch is cited to the cited	npletion of the International Search	"T" later document published after or priority date and not in condicited to understand the princit invention. "X" document of particular releval cannot be considered hovel of involve an inventive step. "Y" document of particular releval cannot be considered to involve document is combined with or ments, such combination being in the art. "4" document member of the same.	itle or theory underlying the ince: the claimed invention or cannot be considered to ince: the claimed invention is an inventive step when the is or more other such docu-jobylous to a person skilled a patent family
			Signature of Authorized Officer	
Internation	nal Searchini) Matriatria	James Ketter	_ ebw

	MTS COMBIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SMEET	l a le cat és Clair No
II DOCUME	NTS CONDIDENCE TO THE PROPERTY OF THE PROPERTY	Reseant to Claim No
tegory •	Citation of Document, with indication, whose appropriate	1-87
-7.	GENE, Volume 44, issued 1986. A.R.	1-0/
	The state of the s	
	oligodeoxynucleotides", pages 177-183,	
	see entire document.	
4.0	PROCEEDINGS OF THE NATIONAL ACADEMY OF	1-87
Y^{-}	GOTENOE VOLUMO 87 15SUEO AUGUST 19907	
	1 "Dentides on Dudue: A	* .
	wast library of Deptides for Identitying	
	ligands". pages 6378-6382. see entire	
	document.	1000
. ter ફ. 48°.	SCIENCE. Volume 249. issued 27 July 1990,	1-87
	t t haviin "Random Peblice Libraries.	
	a course of Specific Protein binging	
	Molecules", pages 404-406, see entire	
	document.	
	SCIENCE. Volume 249. issued 27 July	1-87
1	logo J K Scott, "Searching IOT	
	pentide Ligards with an Epitope Library.	
	pages 386-390, see entire document.	
		1-87
Y	EL. WINNACKER. "From Genes to Clones: Introduction to Gene Technology".	
	Germany), See pages 276-279, especially	
	Table 7-4.	
	SCIENCE. Volume 228. issued 14 June	8.9.24-26
Y	1985. G.P. Smith. "Filamentous Fusion	32-34.
	Phage Novel Expression Vectors That	55-57
	nianiay Cloned Antigens on the Virion	64-66. 73-75.
	Surface". pages 1315-1317. see entire	81-87
	document.	
()		
\		
i .		
		-